

A:Cross-references: GB:AE005174; NID:q12518695; PIDN:AG59017.1; GSPDB:GN0145; UWGP:Z53
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: pldA
C:Superfamily: bacterial phospholipase A1

Query Match 15.1%; Score 357; DB 2; Length 289;
Best Local Similarity 39.3%; Pred. No. 2.4e-21;
Matches 84; Conservative 29; Mismatches 91; Indels 10; Gaps 4;

OY 226 RAPELKFQVSVYVKAEDLMDGSDLMFGYQOQSHMOIFENGNSRPFVHDQPEIFLTQ 285
DB 83 RKDEKRFQSLAFPLMRGILGPNVLSGASYOKSMWOLSNSESSPFRETTEPQLGCF 142
OY 286 PV-YSDLPMDGKVRMIGAVHNSGESAKLSRSNRRAYLMAEMKMLTYVPRITGRIF 344
DB 143 ATDYEFAGM--TLRDYEMGYNHDSNGRSDPTSRSMNRLYTRLMAENGMLVYKRWYV- 199
OY 345 KEGSSQPDNDLDIDYGYDFRFLYOLENKSNSIGVRRNPRSGKALDIDYVPLGK 404
DB 200 -----GNTDNDPDIRKMGYYOLKIGYHL-GDAVLSAKGYNMNTGYGAGELGSLPYTK 253
OY 405 GTSGYFOIFOGYQSLIDYVNEHATSEFGVGLMND 438
DB 254 HVRLTYVSGSGESLIDYNQTRVGVGLMND 287

RESULT 8
C36971
outer membrane phospholipase A (EC 3.1.1.-) precursor - Proteus vulgaris

C:Species: Proteus vulgaris
C:Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 18-Jun-1999
C:Accession: C36971; S40130
R:Brok, R.G.P.M.; Birkman, E.; van Bortel, R.; Bekkers, A.C.A.P.A.; Verheij, H.M.; Tomm
J. Bacteriol. 176, 861-870, 1994
A:Title: Molecular characterization of enterobacterial pldA genes encoding outer membra
A:Reference number: A36971; MUID:94131966
A:Accession: C36971
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-289

A:Cross-references: EMBL:X76902; NID:q436889; PIDN:CAA54224.1; PID:q436890
A:Note: authors translated the codon GAG for residue 74 as Gly and CGA for residue 115 a
C:Genetics:
A:Gene: pldA
C:Superfamily: bacterial phospholipase A1
C:Keywords: carboxylic ester hydrolase

Query Match 13.6%; Score 321.5; DB 2; Length 289;
Best Local Similarity 33.3%; Pred. No. 1.8e-18;
Matches 87; Conservative 42; Mismatches 115; Indels 17; Gaps 7;

OY 179 LDRNTPILMSSSRPHNPMVLPITFMHGKPNRSPTNSHEAKQFTPEFRAPELKPFQSVKV 238
DB 43 LOEHNP-FTLYPVSNLTLYTY-----TSDLNKKAISYMSDNA-NKDEKFLQSLAF 95
OY 239 KAAEDLMGTDSDLMFGYQOQSHMOIFENGNSRPFVHDQPEIFLTQPV-YSDLPMDK 297
DB 96 PLMRILDLNLSLGSASTYORSWOLSNTEGSAFRETTEPQLGCFATDYSVGW--TL 153
OY 298 RMIGGAVHNSGESAKLSRSNRRAYLMAEMKMLTYVPRITGRIFKEGSGSQPDND 357
DB 154 RDAEFGYHNQNGRSDPTSRSMNRLYTRLMAENGMLVYKRWYV-----GDTSDKN 207
OY 358 ILDYGVDVDFLYOLENKSNSIGVRRNPRSGKALDIDYVPLGKISGYFOFGY 417
DB 208 ITKWTGYQLTIGYOL-GEAVLSAKGYNMNTGYGAGELGSLPYTKHVRFTQVSGYG 266
OY 418 OSIDYNHEATSEFGVGLMND 438
DB 267 ESLIDYDENQTRVGVGLMND 287

RESULT 9
C64582
phospholipase A1 precursor - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: C64582
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKe
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467
A:Accession: C64582
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-355 <TOM>

A:Cross-references: GB:AE000564; GB:AE000511; NID:q2313602; PIDN:AAD07564.1; PID:q231

Query Match 10.4%; Score 246.5; DB 2; Length 355;
Best Local Similarity 25.1%; Pred. No. 2.8e-12;
Matches 82; Conservative 41; Mismatches 117; Indels 87; Gaps 11;

OY 157 KDKAQLEYAAKQFTPLSLDFDLDRNTPILMSSSRPHNPMVLPITFMHGKPNRSPTPSHE 216
DB 69 KKYLMMDYLGTFYFLPFYHSF-----TPIQWYHPNINP----- 102
OY 217 AKQFTPEFRAPELKFOVSVYVKAEDLMDGSDLMFGYQOQSHMOIFENGNSRPFVHD 276
DB 103 --YORNEF-----KQSLSFVRVFRHLWKGTLYLATYTDWFOYINDPOSAPRRMN 154
OY 277 YQPEIFLTQPVYSDLPMDGKV---RMIGGAVHNSG--ESAKLSRSNRRAYLMAEMKKN 332
DB 155 FMPELIYYVPI-NKRPFGKIGNSEIIMGMHNSGVGAGACQYPRFK-----EENPENQ 209
OY 333 LTVMPRI-----WGRIFKEGSGSQP-----DDNDP 357
DB 210 FPGQPIVIVKDYNGKQKDVWGGCRSVSAGORPFLRWMEKGLIMVAYWPYVYDOSNPN 269
OY 358 ILDYGVDVDFLY-----OLENKSNSIGVRRNPRSGKALDIDYVPLGKISGYF 410
DB 270 LIDYGYNAKIDYRRGRHHEFLDYDIFTQYWRD--RMHGAFLGYTRINPFVGIYA 327
OY 411 OIFOGYQSLIDYVNEHATSEFGVGLMND 437
DB 328 QMFNGYGDGYEYDFVFSNRIGYGRIN 354

RESULT 10
H71930
probable phospholipase a1 - Helicobacter pylori (strain J99)

A:Variety: Strain J99
C:Species: Helicobacter pylori
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: H71930
R:Alm, R.A.; Ling, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
A:Reference number: A71800; MUID:99120557
A:Accession: H71930
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-355 <ARN>
A:Cross-references: GB:AE001479; GB:AE001439; NID:q4154979; PIDN:AAD06029.1; PID:q415
A:Experimental source: strain J99
C:Genetics:
A:Gene: pldA

Query Match 10.2%; Score 240.5; DB 2; Length 355;
 Best Local Similarity 25.1%; Pred. No. 8.6e-12;
 Matches 83; Conservative 44; Mismatches 109; Indels 95; Gaps 13;

Oy 157 KKDAKOLEYAAKQPTPLSLFSDLDNRNPTLWSSRPHPMYVLPITMGKPKNSPPTPSHE 216
 Db KKLMMMDYLGTFLPEVHSF-----PPIQWYHPNINP----- 102
 Oy 217 AKQFPNFRPARELFQVSVKAAEDLMGTDSDLMFGTQOSHMOIFNGKRSRPFRVD 276
 Db 103 ---YORNEP---KFQISFRVPRFRHILMTGTLYLATQNNMFOIYVDPQSAAPRMKN 154
 Oy 277 YQPELFLTPQVSDLPMDCKV---RMIGKAVHNSNG-ESAKLSRSNMR----- 321
 Db 155 FMPRELIVYPT-NKRPFGKIGNFSEIWMQHISNGVAGACQYPRFKEGRPENQFPQ 213
 Oy 322 -----AYLMAG-----MEMK--NLTVMPRIGRIFKEGSGSDPD- 354
 Db 214 PVIYKDYNGQKDYRWGCGRSVAGNALCFVLWMEKGLKIMVAMPYV-----PYDQ 265
 Oy 355 -NPOLDIYGYGDVAFVLY-----QLENKSNISGTVKRNPSGKALOLDIYVPLGKI 406
 Db 266 SNPOLDIDWYGNGAKIDYRGRHHPLELDYDFTQYWRD--RMHGAFLGYTYRINP 323
 Oy 407 SGYFOIFQYGGSLIDYHNHATSPFGVGLMLN 437
 Db 324 GIVQWFMNGYGDGLXEVDFSNRIGVGRIN 354

RESULT 11
 TYVTRR
 protein kinase (EC 2.7.1.37) raf - rat
 N.Alternate names: kinase-related transforming protein raf; raf proto-oncogene protein-8
 C.Species: Rattus norvegicus (Norway rat)
 C.Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 11-Jun-1999
 C.Accession: B26126
 R.Ishikawa, F.; Nakano, M.; Sugimura, T.
 Mol. Cell. Biol. 7, 1226-1232, 1987
 A.Title: Rat c-raf oncogene activation by a rearrangement that produces a fused protein.
 A.Reference number: A26126; MUID:87172791
 A:Accession: B26126
 A:Molecule type: mRNA
 A:Residues: 1-602 <1SH>
 A:Cross-references: GB:M15428; NID:g206546; PIDN:AAA42002.1; PID:g206547
 C:Genetics:
 A:Gene: raf
 C:Superfamily: rat protein kinase raf; protein kinase homology
 C:Keywords: ATP; autophosphorylation; phosphoprotein; phosphotransferase; serine/threonine
 F:301-567/Domain: protein kinase homology <KIN>
 F:309-317/Region: protein kinase ATP-binding motif
 F:329/Active site: Lys #status predicted
 F:453/Binding site: phosphate (ser) (covalent) (by autophosphorylation) #status predicted

Query Match 4.8%; Score 112.5; DB 1; Length 602;
 Best Local Similarity 19.3%; Pred. No. 0.43;
 Matches 93; Conservative 71; Mismatches 174; Indels 145; Gaps 23;

Oy 23 QQAQVNPVPAFVDEVRENDLQDNELPIDVQSAFQSAFTDANPLDEHPPELYTTALEN 82
 Db 81 ELEKLNQVYKLETKETAEQDRLGI-----QSFTRAKKELEAKRDLIRT---N 131
 Oy 83 KMLMLNSALNODIMRLACVDTLVHGEFPAVITKRSITLDETIWQ--TIKKRPQVIYOE 140
 Db 132 EKLSEVEVLTEDVRL--NEKLKESNT---TKGELQLLDELQASDVVYKREKLEOE 186
 Oy 141 -----FTDPIFLGNENK-----MLTKDAKOLEYAAKOPTPLS 174
 Db 187 KELLNQNQSWLMTLTKTDELALGREGNELLKCTLEKKKEDDAIRSHSEASPSA 246
 Oy 175 LSFDDLRNNTPLMSRPHNPMVLPITMGKPKNSPPTPSHEAKOPTPNEFRAPLEKQV 234
 Db 247 LS-SSPNMLSPYQWSPKRP-----VPAQREBARAFGSGTOEKNKIRPRGQDSVYWEI 298

Oy 235 SVKVAEDLMGTDSDLMFGTQOSH-----QIFNGKNSRPFVHDXOPEI----- 281
 Db 299 EASEVNLSTRISGS---FGTVYKGMHGDVAVKILKVDPPPEQLOAEFNEVAVLRKTR 355
 Oy 282 -----FLTQ-----PYSDU--PMDGKVRMI-----GMGAVHR 307
 Db 356 HYNILLFMGYMTKDNALITWQCESSSLKHLHYQETKQOMQLDIARQAGMDYLAH 415
 Oy 308 SNGESAKLSRSNMRAYLMAGMEWK---NLVMPRIWGRIFKEGSGSO----- 351
 Db 416 KNIHRDMKS--NMLPLHGLTYKIDFGLATVYKSH-----SGSQVEQPTGSVLM 466
 Oy 352 -----PDDNDIL--DYGYGDVAFVLYOLEKSNISGTVKRNPSGKALOLDIYVPL 402
 Db 467 APEVIRMODNPPSPFQSDVSYGIV--LYEL-----MTGELPYSHINRDQI---IFWV 515
 Oy 403 GRG 405
 Db 516 GRG 518

RESULT 12
 A29003
 cellulase (EC 3.2.1.4), alkaline - Bacillus sp.
 N.Alternate names: endo-1,4-beta-glucanase
 C.Species: Bacillus sp.
 C.Date: 21-May-1988 #sequence_revision 21-May-1988 #text_change 21-Jul-2000
 C.Accession: A29003
 R.Fukumori, F.; Kudo, T.; Narahashi, Y.; Horikoshi, K.
 J. Gen. Microbiol. 132, 2329-2335, 1986
 A.Title: Molecular cloning and nucleotide sequence of the alkaline cellulase gene fro
 A.Reference number: A29003; MUID:87085443
 A:Accession: A29003
 A:Molecule type: DNA
 A:Residues: 1-800 <FUR>
 A:Cross-references: GB:D00066; GB:N00066; NID:g216223; PIDN:BAA00045.1; PID:g216224
 A:Experimental source: strain 1139
 C:Function: A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
 A:Pathway: cellulose degradation
 C:Superfamily: Bacillus sp. alkaline cellulase; Thermotoga xylanase A amino-terminal
 C:Keywords: glycosidase; hydrolysis; polysaccharide degradation
 F:585-726/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA>

Query Match 4.7%; Score 110.5; DB 2; Length 800;
 Best Local Similarity 20.0%; Pred. No. 0.96;
 Matches 96; Conservative 53; Mismatches 163; Indels 169; Gaps 25;

Oy 9 TILSLCFALIAQQAQVNPVPAFVDEVREND---LQDNELPIDVQSAFQSAFTDRA 65
 Db 7 TQOLISSILILV-LTSLPPTALAAEGNTRDEFKHLGLNDVKKRPSGALQLOEVDQ 65
 Oy 66 NPL-DEHEPELY-----TTALENKTMLINCSAL-----NODIMLACY-DTIVHGETP 111
 Db 66 MTLVQDHGKGIQLRGSMSTIGLOMFPEIINDNAKALANDMESMITLAVYGENGASPN 125
 Oy 112 AVIKTKRSIRLDETT-----WQT-----IKKRPQVIY 138
 Db 126 ELIKSHVIGIDILAIENDYVIVDMVHAPGDRDPAVGAEDFFRIDALYPPNNHITY 185
 Oy 139 QETDP-----IFLMGMEKMLTKKDAKOLEYAAKOPTPLSSTDLDNR---NTPLW 187
 Db 186 ELANEPSSNNNGAGIPNNEEGNAV---EYADPIEMLRDSGNADNIIIVGSPNW 240
 Oy 188 SSRP-----HNPMVLPILF--MHGKPKNS--PNTPSHEAKOPTPNEFRAPLEKQ 233
 Db 241 SQRPDLAANPIDDHTMTTIVHYTGSNAASTESYPERFPNBERGVMNTRYA---LE 296
 Oy 234 VSVKVAEDLMGT-----DSDLWFGYIQO-----SHWOIFNGKNS---RP 271
 Db 297 NGVAVATE--WQTSQANGDGPYFDEADVWIEFLNENNISWANWISITN-KNEVSGAFYP 353

[illegible]

RESULT 13

JT0611

cellulase (EC 3.2.1.4), alkaline - *Bacillus* sp. (strain KSM-64)

N:Alternate names: endo-1,4-beta-glucanase

C:Species: *Bacillus* sp.

C:Date: 09-Oct-1992 #sequence_revision 09-Oct-1992 #text_change 10-Dec-1999

C:Accession: JT0611

R:Sumitomo, N.; Ozaki, K.; Kawai, S.; Ito, S.

R:Biosci. Biotechnol. 56, 872-877, 1992

A:Title: Nucleotide sequence of the gene for an alkaline endoglucanase from an alkalophilic

A:Reference number: JT0611, MUID:92305459

A:Accession: JT0611

A:Molecule type: DNA

A:Residues: 1-822 <SOM>

A:Cross-references: GB:M84963; NID:g289264; PIDN:AAA73189.1; PID:g289266

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel-

A:Pathway: cellulose degradation

C:Superfamily: *Bacillus* sp. alkaline cellulase; *Thermotoga* xylanase A amino-terminal repeat

C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation

C:585-726/Domain: *Bacillus* xylanase A amino-terminal repeat homology <TXN>

Query Match	4.7%	Score	110.5	DB 2	Length	822			
Best Local Similarity	20.0%	Pred. No.	1						
Matches	96	Conservative	53	Mismatches	163	Indels	169	Gaps	25

OY		9	TLSLISCAALIAQQAQAVPVPVAFVDVEBSEND-----LGSDNELPIDVGSATGASHTDA	65
Dd		7	TKOLISSLLIIVL-LLSLFPPALABEGNTRFEDNKHLIGNDYNKRRSEAGHLOIQEYDGO	65
OY		66	NPL-DEHEPELY-----TTALENTMLDINGSAL-----NODIMRLACY-DPLVHGSTR	111
Dd		66	MVLVDHGEKLTOLRGMSHTGLQMPPELLINDMAYAKALANDMESNMIRILAMVVGENGYSNP	1252
OY		112	AVITKKSIRIDEII-----WQT-----IKGRPVITY	138
Dd		126	ELISKRYATKGIDLIENDMYIYDMVHAHPGDPRPYAGAEDFRDIALYPPNNPIIY	185
OY		139	QETDP-----TEFLGNNEGMLTKRKDAKOLEYAAROKFTPLSLSPDLRN-----NPLW	187
Dd		186	ELANEPSSNNNGAGCIPINNEGGMAVAK-----EXADRIVELLRSGNADDNIIVGSPNW	240
OY		188	SSRP-----HNPNVYLPT--MHGRPNNS--PNTSSEAHOFTPNEFRAPELKFO	233
Dd		241	SQRDLAANDPIDDHHTMTVHFYTGSAASTEEYPREPTEPSEGNVMSTRVA-----LE	266
OY		234	VSVYVKAAMDLMGT-----DSDLMPFGVTQQ-----SHMOIFENSKNS-----RP	271
Dd		297	NGVAVFATE--WGTSQANGDGGRPFEDADWVIEFLNENNISMAMWSLTLN-KNEYSGAFTP	353
OY		272	FRVH-----DYQELTELTPQVYSDLPW-----D	294
Dd		354	FELCKSNATSLDPRPDQVWPBEELSLSGEYVARIKGVNERP---IDRTKYTFKLAMPFD	410
OY		295	GKVAYIMGAVHHNGES-----AKLSRBNRAYLLAGMEKNLTLYMPRIJG	341
Dd		411	GTKQGFGV-----NGSPVEDVVITENEAGALKTIGSLDANSVSEGVMYMANARSLADGWG	464

QY	342 R 342
Db	465 K 465

RESULT 14

JC7532

cellulase (EC 3.2.1.4), alkaline - *Bacillus* sp. (strain KSM-S237)

N:Alternate names: endo-1,4-beta-glucanase

C:Species: *Bacillus* sp. (strain KSM-S237)

C:Date: 31-Mar-2001 #sequence_revision 31-Mar-2001 #text_change 31-Mar-2001

A:Accession: JC7532; PC7107

R:Hakamada, Y.; Hataida, Y.; Koike, K.; Yoshimatsu, T.; Kawai, S.; Kobayashi, T.; Ito, Biosci. Biotechnol. Biochem. 64, 2281-2289, 2000

A:Title: Deduced amino acid sequence and possible catalytic residues of a thermostably

A:Reference number: JC7532; MUID:21036886

A:Accession: JC7532

A:Molecule type: DNA

A:Residues: 1-824 <HAK>

A:Cross-references: DDBJ:AB018420

A:Experimental source: strain KSM-S237

A:Accession: PC7107

A:Molecule type: protein

A:Residues: 31-50 <HA2>

C:Comment: This enzyme is a thermostable, alkaline enzyme used as effective detergent

C:Genetics:

A:Gene: Egl-237

C:Keywords: hydrolase; glycosidase

Query Match	4.68;	Score 109;	DB 3;	Length 824;
Best Local Similarity	20.08;	Pred. No. 1.3;		
Matches	96;	Conservative	59;	Mismatches 159;
				Indels 166;
				Gaps 28;

OY	9	TLISLISFCALIAIQAQAQAVNPAPAFDEVRESD---	LCODNDELPIDVQSAPOSASTRA	65
Db	7	TKOLISLILVL--LISLPAALAAAGNTREDFKILLNDVKRPS	EAGALQLOEVDQ	65
OY	66	NPL-DEHEPELY-----TTALENKTMLINCSAL-----	NODIMRLACDYTLVHGE---	109
Db	66	MTLVDQGEKTIQLRGMTHTGLQMFPELINDMVKALISND	MSMIRLAMY-----VGENY	121
OY	110	--PRAVIKTK-----RSIRLEDTI---WOT-----	-----IKGP	134
Db	122	ATNPBLIKQAVIDGIELIALIENDMYVIVDMVHVAR	GDPRDPVYAGAKDFPREIALYLPNPP	181
OY	135	QVIVQETTD-----IFLMGKMKLTKKQAKQLEVA	AKQFTPLISFDLDRN-----N	183
Db	182	HLIETLANEPSSNNNGAGAPINNEEEMKAVK---	EYADPIVEIMIRKSGNADNNITIVG	236
OY	184	TPPLSSRP-----NHPMYVLPIFMHGKPNRSPNP	PSHEAKQFTPNFERABELK--	231
Db	237	SPNNSQRPDLAADNPIDDDHTTWTYTHFYGSHAAP	ESYPSR-----PNSRGVMMSMT	291
OY	232	---FQSVVKYKAADLMGT-----DSDLMEFYTQ---	SHWJFNGKNS--	269
Db	292	RYALENGVAVFATE--WGTSGAAGDGGPYFDEAD	VAWIEFLANNNISWAMMSLTN-KNEYS	348
OY	270	---RPRFV-----HDYOB-ELFIET-----	QPY-----YSQSLP--	293
Db	349	GALTFPEELGASNAITNLDPGRDHMAAEBELSL	GEVYRARIKGVNTEPIDRTKYTVLWDF	408
OY	294	-DGKVRIRIGAGVAHHSNGESAKLSRSVMNRY	ALMAGNE-----WKNLVMPRIAGR	342
Db	409	NDGIRKQ--GRGVMSDSPNKNELIAVDNENNT	LAKVSGLDVSNVSDGPFMANARLASANGMK	466

RESULT 15
 T20122
 hypothetical protein P25B3.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
 :Accession: T20122; T21324

R. McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19225
A:Accession: T20122
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-901 <NTL>
A:Cross-references: EMBL:Z70750; PIDN:CAA94750.1; GSPDB:GN00023; CESP:F25B3.1
A:Experimental source: clone C50F4
R.Gardner, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19406
A:Accession: T21324
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-901 <NT>
A:Cross-references: EMBL:Z70752; PIDN:CAA94758.1; GSPDB:GN00023; CESP:F25B3.1
A:Experimental source: clone F25B3
C:Genetics:
A:Gene: CESP:F25B3.1
A:Map position: 5
A:Intons: 25/3; 47/2; 199/1; 242/1; 315/2; 544/3; 600/1; 626/1; 646/1; 666/3; 706/3; 75

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:18:34 ; Search time 44.45 seconds
(without alignments)
757,461 Million cell updates/sec

Title: US-09-787-083-6

Perfect score: 2360

Sequence: 1 MKVSLSTLTLISCFALIA.....YNHEATSRGVLMLNDWGL 442

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR_68:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	815	34.5	382	2	H81195 phospholipase A1, probable phospholip
2	815	34.5	409	2	H81831 outer membrane pho
3	373	15.8	286	2	B36571 phospholipase A1 (
4	366	15.5	329	2	D81279 outer membrane pho
5	358	15.2	289	2	A36971 PSECA1
6	357	15.1	289	1	E86059 outer membrane pho
7	357	15.1	289	2	C36971 outer membrane pho
8	321.5	13.6	289	2	C64582 phospholipase A1 p
9	246.5	10.4	355	2	H71930 protein kinase (EC
10	110.5	4.7	602	1	TVTRRR cellulase (EC 3.2.
11	109.5	4.6	800	2	A29003 cellulase (EC 3.2.
12	109.5	4.6	822	2	JT0611 cellulase (EC 3.2.
13	108	4.6	822	2	JC7532 cellulase (EC 3.2.
14	105.5	4.5	901	2	T20122 hypothetical prote
15	103	4.4	783	2	JC5467 cellulase (EC 3.2.
16	101	4.3	719	2	A42893 penicillin-binding
17	101	4.3	4307	2	T42642 hypothetical prote
18	101	4.3	4307	2	T42642 hypothetical prote
19	99.5	4.2	1658	2	S28031 penicillin-binding
20	99	4.2	719	2	S28031 penicillin-binding
21	99	4.2	719	2	S28033 penicillin-binding
22	99	4.2	5005	2	F82884 hypothetical prote
23	98.5	4.2	660	2	E83656 methionyl-L-cRNA syn
24	98.5	4.2	810	2	A49744 AMP deaminase (EC
25	98	4.2	719	2	S28034 penicillin-binding
26	98	4.2	719	2	S28032 penicillin-binding
27	97.5	4.1	1478	2	S20117 protein kinase BCK
28	96.5	4.1	797	2	T46737 X-Pro dipeptidyl-P
29	96.5	4.1	838	2	A96557 probable receptor

30	96	4.1	608	2	S28036 penicillin-binding
31	95	4.0	719	2	S28035 penicillin-binding
32	95	4.0	719	2	S28038 penicillin-binding
33	95	4.0	719	2	S28037 penicillin-binding
34	95	4.0	765	2	T35719 cellulase - Strept
35	94.5	4.0	397	2	A35136 cellulase (EC 3.2.
36	94.5	4.0	601	2	T26062 hypothetical prote
37	94.5	4.0	655	2	T26061 hypothetical prote
38	94	4.0	1115	2	D69517 DNA polymerase III
39	93.5	4.0	564	2	T40777 ferric reductase t
40	93	3.9	791	2	H96839 S-receptor kinase
41	92.5	3.9	857	1	A41369 hepatocyte growth
42	92.5	3.9	1379	1	S01254 hypothetical prote
43	92	3.9	324	2	B69521 AMP deaminase homo
44	92	3.9	888	2	S50801 gene 14 protein -
45	91.5	3.9	486	2	S30959

ALIGNMENTS

Query Match	Best Local Similarity	Score	DB 2:	Length	382:
Matches 163;	Conservative 67;	Mismatches 111;	Indels 24;	Gaps 8;	
QY 87	INCSALNODIMRLACYPDLVHGETPAVY-----KTKRSIRLDETIMQTI-KGKPOVYQE	140			
DB 33	LOCALALDNTVTRLACVIRFAOOLPSSAGGEGSKAVLNITEVRSLSKGEAVIYER	92			
QY 141	TTDPFILMNEKGMLTRKKAQOLEYAAKOPTPLSLSFEDLRNN-TPLMSSPHNPVYLP	199			
DB 93	GDAL-----PADSAGETADITYPLSLMVDLKDNDLRLGLVGEHNPVYIMP	139			
QY 200	IFPHGKRNRSPTNPFSHAR-QFTNEFRAPELKQVSVYKKAADMDLGTSDMLFGTQQ	258			
DB 140	LWYNNSPVAPGSPTRCTTQVEKFGQKRAETKIQVFSKSLADLDLKTTRADLFEVGTOR	199			
QY 259	SHMOIFR-GKNSRPFRRHDYOPETFLTOPVYSDLPWVGKVRMIGAVHNSNGESATLSR	317			
DB 200	SDMOIYOGKRSAPFRNTDYKPELFLOPYKADLPFGGRMLAGVYHOSNGOSRPSR	259			
QY 318	SWNRAYLMAEMKNTLVMPRIMGRIFKESGSGQPDNDPILDYGYGDVRFYOLENKS	377			
DB 260	SWNRIVYAMAGMEKGLVIVRVWVAFDQ-SGDR-NDNPDIADYMGVGDVLYQRLNDRQ	317			
QY 378	NISGTVYRNRSKGALQLDYVYPLGKIGISGYROIPOGSGOSLIDYVHEATSRGVGLMLN	437			
DB 318	NVYSVLRNPKRTGALAEAYTPPIKGLKGVYVGFHGESESLIDYVHNKONGIGIQLMFN	377			
QY 438	DWNGGL 442				

Db 378 DLDGI 382

RESULT 2

Probable phospholipase NMA2021 [Imported] - *Neisseria meningitidis* (strain Z2491 serogroup C)
C.Species: *Neisseria meningitidis*
C.Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C.Accession: H61831
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
: Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream
Nature 404, 502-506, 2000
A.Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.
A.Reference number: A81775; MUID:20222556
A.Accession: H61831
A.Status: preliminary
A:Molecule type: DNA
A:Residues: 1-409 <PAR>
A:Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CA85240.1; PID:g738065
C.Experimental source: serogroup A, strain Z2491
C.Genetics:
C:Gene: NMA2021

Query Match	34.58;	Score 815;	DB 2;	Length 409;
Best Local Similarity	44.78;	Pred. No. 2.1e-58;		
Matches 163;	Conservative 67;	Mismatches 111;	Indels 24;	Gaps 8

[illegible]

RESULT 3

outer membrane phospholipase A (EC 3.1.1.-) precursor - *Klebsiella pneumoniae*
C:Species: *Klebsiella pneumoniae*
C:Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 18-Jun-1999
C:Accession: B36971; S40129
R:Brok, R.G.P.M.; Brinkman, E.; van Boxtel, R.; Bekkers, A.C.A.P.A.; Verheij, H.M.; Tomm
J. Bacteriol. 176, 861-870, 1994
A:Title: Molecular characterization of enterobacterial *pldA* genes encoding outer membra
A:Reference number: A36971; MUID:94131966
A:Accession: B36971
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-286 <BRO>
A:Cross-references: EMBL:X75901; NID:9436880; PTDN:CAA54223.1; PID:9436881

A:Note: authors translated the codon AAG for residue 112 as Arg
C:genetics:
A:Gene: plda
C:Superfamily: bacterial phospholipase A1
C:Keywords: carboxylic ester hydrolase

Query Match	15.8%;	Score 373;	DB 2;	Length 286;
Best Local Similarity	42.1%;	Pred. No. 9.6e-23;		
Matches 90;	Conservative 27;	Mismatches 87;	Indels 10;	Gaps 5;

QY	226	RAPLEKFEVSYKVAABEDLMGTDSDLMGYQOOSIMQJFNCKNSRPFVHYOPEIFLQ	285
	1	:	
	2	:	
	3	:	
	4	:	
	5	:	
	6	:	
	7	:	
	8	:	
	9	:	
	10	:	
	11	:	
	12	:	
	13	:	
	14	:	
	15	:	
	16	:	
	17	:	
	18	:	
	19	:	
	20	:	
	21	:	
	22	:	
	23	:	
	24	:	
	25	:	
	26	:	
	27	:	
	28	:	
	29	:	
	30	:	
	31	:	
	32	:	
	33	:	
	34	:	
	35	:	
	36	:	
	37	:	
	38	:	
	39	:	
	40	:	
	41	:	
	42	:	
	43	:	
	44	:	
	45	:	
	46	:	
	47	:	
	48	:	
	49	:	
	50	:	
	51	:	
	52	:	
	53	:	
	54	:	
	55	:	
	56	:	
	57	:	
	58	:	
	59	:	
	60	:	
	61	:	
	62	:	
	63	:	
	64	:	
	65	:	
	66	:	
	67	:	
	68	:	
	69	:	
	70	:	
	71	:	
	72	:	
	73	:	
	74	:	
	75	:	
	76	:	
	77	:	
	78	:	
	79	:	
	80	:	
	81	:	
	82	:	
	83	:	
	84	:	
	85	:	
	86	:	
	87	:	
	88	:	
	89	:	
	90	:	
	91	:	
	92	:	
	93	:	
	94	:	
	95	:	
	96	:	
	97	:	
	98	:	
	99	:	
	100	:	
	101	:	
	102	:	
	103	:	
	104	:	
	105	:	
	106	:	
	107	:	
	108	:	
	109	:	
	110	:	
	111	:	
	112	:	
	113	:	
	114	:	
	115	:	
	116	:	
	117	:	
	118	:	
	119	:	
	120	:	
	121	:	
	122	:	
	123	:	
	124	:	
	125	:	
	126	:	
	127	:	
	128	:	
	129	:	
	130	:	
	131	:	
	132	:	
	133	:	
	134	:	
	135	:	
	136	:	
	137	:	
	138	:	
	139	:	
	140	:	
	141	:	
	142	:	
	143		

RESULT 4

Phospholipase 1 (EC 3.1.1.32) Cj1351 [imported] - Campylobacter jejuni (strain NCTC C-Spectes: Campylobacter jejuni)
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 18-Aug-2000
C:Accession: DB1279
R:Patkhil, J., Wren, B.W., Mungall, K., Ketley, J.M., Churcher, C., Basham, D., Chillingworth, F., Quail, M., Raddadream, M.A., Rutherford, K.M., VanVleet, A., Whitehead, S., Baer, C.W., 4031, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: A81250; MUID:20150912
A:Accession: DB1279
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-329 >PAR>
A:Cross-references: GB:AL139078; GB:AL111168; NID:65968723; PIDN:CA873778.1; PID:659686
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: pldA; Cj1351
C:Keywords: carboxylic ester hydrolase

Query Match	15.5%;	Score 366;	DB 2;	Length 329;
Best Local Similarity	33.2%;	Pred. No. 4.4e-22;		
Matches 105;	Conservative 48;	Mismatches 125;	Indels 38;	Gaps 13;

[illegible]

OY 295 GAVRMIGCAVHHSNGES-----AKLSRSMRAVILMAGMEKNLTWMPRIWG 341
 Db 411 GTRKQGFV-----NDSPEDVVIENEGALKLGLDASNDVSEGNWYANRLSADGWG 464
 OY 342 R 342
 Db 465 K 465

RESULT 13
 J70611
 cellulase (EC 3.2.1.4), alkaline - Bacillus sp. (strain KSM-64)

N.Alternate names: endo-1,4-beta-glucanase
 C.Species: Bacillus sp.
 C.Date: 09-Oct-1992 #sequence_revision 09-Oct-1992 #text_change 10-Dec-1999
 C.Accession: J70611
 R.Sunltomo, N.; Ozaki, K.; Kawai, S.; Ito, S.
 Biosci. Biotechnol. Biochem. 56, 872-877, 1992
 A.Title: Nucleotide sequence of the gene for an alkaline endoglucanase from an alkalophilic
 A.Reference number: J70611; MUID:92305459
 A.Accession: J70611
 A.Molecule type: DNA
 A.Residues: 1-822 <SD>
 A.Cross-references: GB:M84963; NID:9289264; PID:NAA73189.1; PID:9289266
 C.Function: A.Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
 A.Pathway: cellulose degradation
 C.Superfamily: Bacillus sp. alkaline cellulase; Thermocoga xylanase A amino-terminal ref
 C.Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
 F:585726/Domain: Thermocoga xylanase A amino-terminal repeat homology <TX>

Query Match 4.6%; Score 109.5; DB 2; Length 822;
 Best Local Similarity 19.8%; Pred. No. 1.2;
 Matches 95; Conservative 54; Mismatches 163; Indels 169; Gaps 25;

OY 9 TLTSLSCFAIIAQAKAPNVAFVDEVRSEND---LGQNDHELPIDVQSAQSASTORA 65
 Db 7 TKQLISSITLVL-LLSLFPALAAEGNTREDNFKHLGNDVKKRSEGAQLQLEVDQ 65
 OY 66 NPL-DEHEPELY-----TTALENKTMLINCSAL-----NODIMRLACY-DTLVHGERT 111
 Db 66 MTLVDQHGKIDLRGMSTHGLQWPEIILNDNAYKALANDMESNMIRLAMYGENGYASNP 125
 OY 112 AVIKTKRSIRIDET-----WQT-----IKGRPOVY 138
 Db 126 ELIKSRVIGIDLALENDMYIVDWHVHAPGDRPVYAGADEFRDIALALYPPNPHIT 185
 OY 139 QETTPD-----IFLMGNEKGMILTKKAKOLEYAAKQFTPLSLSFDDLRN-----NTPLW 187
 Db 186 ELANEPSSNNNGAGIIPNNEEGMNAVK-----ETADPIVEMLRDSGNADNDNIIIVGSPW 240
 OY 188 SSRP-----HNPMYVLPF--MHGKPNRS--PNTPSHEARQFTPNFRAPELKQ 233
 Db 241 SQRPLADNPIDIDHTHTYVHFYTGSHAASTESTPPEIPNBERGNVMSNTYA-----LE 296
 OY 234 VSVKKAADLMGT-----DSDLMFGYTQO-----SHWQIFNGKNS-----RP 271
 Db 297 NCVAVFATE--WGTSQANGDGPGYFDEADVWIEFLNENNISMANNSLTN-KNEVSGAFTP 353
 OY 272 FRVH-----DYQPIFLTPQVYSDLPW---D 294
 Db 354 FELGKSNATSLDPPGPDQVWVPEELSLSGEYVRAIRIKGVNTEP---IDRTKTKVLMDFND 410
 OY 295 GAVRMIGCAVHHSNGES-----AKLSRSMRAVILMAGMEKNLTWMPRIWG 341
 Db 411 GTRKQGFV-----NDSPEDVVIENEGALKLGLDASNDVSEGNWYANRLSADGWG 464
 OY 342 R 342
 Db 465 K 465

RESULT 14
 JC7532
 cellulase (EC 3.2.1.4), alkaline - Bacillus sp. (strain KSM-S237)

N.Alternate names: endo-1,4-beta-glucanase
 C.Species: Bacillus sp. (strain KSM-S237)
 C.Date: 31-Mar-2001 #sequence_revision 31-Mar-2001 #text_change 31-Mar-2001
 C.Accession: JC7532; PC7107
 R.Hakamada, Y.; Hatada, Y.; Kolke, K.; Yoshimatsu, T.; Kawai, S.; Kobayashi, T.; Ito,
 Biosci. Biotechnol. Biochem. 64, 2281-2289, 2000
 A.Title: Deduced amino acid sequence and possible catalytic residues of a thermostabl
 A.Reference number: JC7532; MUID:21036886
 A.Accession: JC7532
 A.Molecule type: DNA
 A.Residues: 1-824 <HA>
 A.Cross-references: DDBJ:AB018420
 A.Experimental source: strain KSM-S237
 A.Accession: PC7107
 A.Molecule type: protein
 A.Residues: 31-50 <HA>
 C.Comment: This enzyme is a thermostable, alkaline enzyme used as effective detergent
 C.Genetics:
 A.Gene: Egl-237
 C.Keywords: hydrolase; glycosidase

Query Match 4.6%; Score 108; DB 3; Length 824;
 Best Local Similarity 19.8%; Pred. No. 1.5;
 Matches 95; Conservative 60; Mismatches 159; Indels 166; Gaps 28;

OY 9 TLTSLSCFAIIAQAKAPNVAFVDEVRSEND---LGQNDHELPIDVQSAQSASTORA 65
 Db 7 TKQLISSITLVL-LLSLFPALAAEGNTREDNFKHLGNDVKKRSEGAQLQLEVDQ 65
 OY 66 NPL-DEHEPELY-----TTALENKTMLINCSAL-----NODIMRLACYDTLVHGE--- 109
 Db 66 MTLVDQHGKIDLRGMSTHGLQWPEIILNDNAYKALANDMDSNMIRLAMY-----VGENGY 121
 OY 110 --TPAVIKTK-----RSIRIDET---WQT-----IKGRP 134
 Db 122 ATNPelikQRIYDGLIELAIENDMYIVDWHVHAPGDRPVYAGAKDFREIRIALYPPNP 181
 OY 135 QVYQETTPD-----IFLMGNEKGMILTKKAKOLEYAAKQFTPLSLSFDDLRN-----N 183
 Db 182 HTIYELANEPSSNNNGAGIIPNNEEGMNAVK-----ETADPIVEMLRDSGNADNDNIIIVG 236
 OY 184 TPLMSSRP-----HNPMYVLPFIMHGKPNRSPNTPPSHEARQFTPNFRAPELK-- 231
 Db 237 SPNWSQRDLADNPIDIDHTHTYVHFYTGSHAASTESTPSE-----TPNBERGNVMSNT 291
 OY 232 --FOVSVKKAADLMGT-----DSDLMFGYTQO-----SHWQIFNGKNS-- 269
 Db 292 RYALENGVAVFATE--WGTSQASGDGPGYFDEADVWIEFLNENNISMANNSLTN-KNEVS 348
 OY 270 ---RFRV-----HDYQPIFLTPQVYSDLPW--- 293
 Db 349 GATTFPELGKSNATNLDPGPDQVWVPEELSLSGEYVRAIRIKGVNTEPDRKTKVLMDF 408
 OY 294 -DGKVRMIGCAVHHSNGESAALSRSMRAVILMAGMEKNLTWMPRIWG 342
 Db 409 NDGTRQ--GFGVNSDSPKKEILAVDNNENTLKVSGLDVSNDVSDGNFANARLSANGCK 466

RESULT 15
 T20122

hypochemical protein F25B3.1 - Caenorhabditis elegans
 C.Species: Caenorhabditis elegans
 C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
 C.Accession: T20122; T21324
 R:McMurray, A.
 submitted to the EMBL data library, April 1996
 A.Reference number: Z19225
 A.Accession: T20122

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-901 <W1L>
A:Cross-references: EMBL:270750; PIDN:CAA94750.1; GSPDB:GN00023; CESP:F25B3.1
A:Experimental source: clone C50F4
R:Gardner, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19406
A:Accession: T21324
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-901 <W12>
A:Cross-references: EMBL:270752; PIDN:CAA94758.1; GSPDB:GN00023; CESP:F25B3.1
A:Experimental source: clone F25B3
C:Genetics:
A:Gene: CESP:F25B3.1
A:Map position: 5
A:Introns: 25/3; 47/2; 199/1; 242/1; 315/2; 544/3; 600/1; 626/1; 646/1; 666/3; 706/3; 75

Query Match 4.5%; Score 105.5; DB 2; Length 901;
Best Local Similarity 19.6%; Pred. No. 2.8;
Matches 74; Conservative 60; Mismatches 174; Indels 69; Gaps 15;

QY 31 VAFVDEVRSENDLGO-----DNELPIDVQASATQASOTDANPLDEHEPELYTTALENK 83
DB 468 VAMITEIRNOKDLEAVDVHRIPEVPTVQLASRNPALNOPTDADDDAETSNNRPFERS 527
QY 84 TMLI-----NCSALNODIMRLACYDTLVHGETPAVITKRSIRLDETIWQT 129
DB 528 NVSITWVTPGVCAIRASNRASPSKRDELROARDLIEKSTTPATAPNSRKASDEERRREE 587
QY 130 I-----KGPQVYVQETTPITFLMGNEKMLTKKAKOLEYAKOF-----TPLSLSD 178
DB 588 VRRLNEKQHPRTAIPSTSSPYPTFRRIDGNT--DLRIELDVHKKRKPSPPTLVKQ 645
QY 179 LDRNNTPLMSSRPHNPVYLPITFMHGKPNRSPN--TPSHEA--ROFTNEFPAPELKF 232
DB 646 YDPNDI-----PH-----VPAIGRGRTNGNRDSSITPSSASTFDKRYKRGSMRSALKE 694
QY 233 QVSVKVK---AAEDLMGTDSDLMFGYTQO--SHWQ---IFNGKNSRPPRVHDYQPEIF 282
DB 695 SLQLMAKQYGMGNDESSODALATPTKFKFSQWKEKVDVDEGTANELVRIDERISDIT 754
QY 283 LTPQPVYSDLPWMDKVRIGAGVHNSGESAKISRSNNRAYLMAGMWKV--LIVMPRIWG 341
DB 755 AQADVITD-----KIRETEVGSSEEMLTASYLELTNERNTLVHROEYVNIIFTIRQVTS 809
QY 342 RIFKEGS--GSOPDDNP 356
DB 810 EIDQLGKQINEVPDDFP 826

Search completed: November 30, 2001, 14:18:35
Job time: 257 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:27:03 ; Search time 28.02 Seconds

(without alignments)
578.367 Million cell updates/sec

Title: US-09-787-083-6
Perfect score: 2360
Sequence: 1 MKVSLSTLTLSLSCFAILA.....YNHEATSPGVGLMLNDMMGL 442

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	373	15.8	286	1	P37446 klebsiella
2	358	15.2	289	1	P37442 salmonella
3	357	15.1	289	1	P00631 escherichia
4	321.5	13.6	289	1	P37447 proteus vul
5	109.5	4.6	800	1	P06564 bacillus sp
6	98.5	4.2	810	1	P15274 saccharomyc
7	97.5	4.1	1478	1	001389 saccharomyc
8	95	4.0	719	1	004707 streptococc
9	94.5	4.0	397	1	P23548 paenibacill
10	94	4.0	716	1	P97303 mus musculu
11	94	4.0	1115	1	034623 bacillus su
12	92.5	3.9	849	1	009092 brassica ol
13	92.5	3.9	1379	1	P16056 mus musculu
14	92	3.9	324	1	028112 archaeoglob
15	92	3.9	888	1	P40361 saccharomyc
16	91.5	3.9	486	1	005220 mycobacteri
17	91.5	3.9	969	1	055242 streptococc
18	91	3.9	496	1	077229 dictyostell
19	90.5	3.8	467	1	P48997 mus musculu
20	90.5	3.8	1024	1	P47321 mycoplasma
21	90	3.8	1158	1	P48552 homo sapien
22	90	3.8	1788	1	009221 caenorhabdi
23	89	3.8	521	1	P06832 bacillus am
24	88.5	3.8	825	1	P19570 bacillus sp
25	88.5	3.8	1350	1	P08351 homo sapien
26	88	3.7	483	1	084271 chlamydia t
27	88	3.7	537	1	P06716 mus musculu
28	88	3.7	1382	1	P97523 rattus norv
29	87.5	3.7	478	1	P13650 acinetobact
30	87.5	3.7	503	1	P26051 rattus norv
31	87.5	3.7	669	1	P29957 alternomonas
32	87	3.7	637	1	000573 streptococc
33	87	3.7	666	1	088807 rattus norv

34	87	3.7	828	1	BGAL_BRAOL	P49676 brassica ol
35	87	3.7	842	1	LPFC_SALTY	P43662 salmonella
36	87	3.7	917	1	YGJ3_YEAST	P53148 saccharomyc
37	87	3.7	1474	1	A2MG_HUMAN	P01023 homo sapien
38	86	3.6	1008	1	SN14_YEAST	P36048 saccharomyc
39	86	3.6	1087	1	XYNX_CLODM	P38535 clostridium
40	85.5	3.6	353	1	DCUP_BACSU	P32395 bacillus su
41	85.5	3.6	790	1	SEIL_MOUSE	092246 mus musculu
42	85.5	3.6	794	1	SEIL_HUMAN	09ub2 homo sapien
43	85.5	3.6	1374	1	KC9A_SCHRO	009884 schizosacch
44	85	3.6	491	1	TY3H_PNASP	P11982 phasianidae
45	85	3.6	525	1	MP11_RAT	P48965 rattus norv

ALIGNMENTS

RESULT	1	STANDARD	PRT	286 AA.
PAL_KLEPN				
ID	PAL_KLEPN			
AC	P37446:			
DT	01-OCT-1994 (Rel. 30, Last sequence update)			
DT	01-OCT-1994 (Rel. 30, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DETERGENT-RESISTANT			
DE	PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-			
DE	ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).			
CN	PLA.			
OS	Klebsiella pneumoniae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Klebsiella.			
OX	NCBI_Taxid=573;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94131966; PubMed=8300539;			
RA	Brok R.G.P.M., Brinkman E., van Boxtel R., Bekkers A.C.A.P.,			
RT	Verheij H.M., Tommensen J.;			
RT	"Molecular characterization of enterobacterial plda genes encoding			
RT	outer membrane phospholipase A.";			
RL	J. Bacteriol. 176:861-870(1994)			
CC	- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE			
CC	A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.			
CC	- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-			
CC	PHOSPHOCHOLINE + A FATTY ACID ANION.			
CC	- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-			
CC	PHOSPHOCHOLINE + A FATTY ACID ANION.			
CC	- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.			
CC	- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES			
CC	LOCATED THERE.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: X76901; CA54223.1; -.			
DR	PIR: B36971; B36971.			
DR	PIR: S40129; S40129.			
DR	InterPro: IPR003187; PLA1.			
DR	Pfam: PF02253; PLA1; 1.			
KW	Hydrolase; Lipid degradation; Outer membrane; Signal; Calcium.			
FT	SIGNAL	1	20	BY SIMILARITY.
FT	CHAIN	21	286	PHOSPHOLIPASE A1.
FT	ACT SITE	161	161	BY SIMILARITY.
SO	SEQUENCE	286 AA;	32544 MW;	3E39F863085108A3 CRC64;

Query Match 15.8%; Score 373; DB 1; Length 286;
Best local Similarity 42.1%; Pred. No. 5.0e-23;
Matches 90; Conservative 27; Mismatches 87; Indels 10; Gaps 5;

[illegible]

Best Local Similarity 39.3%, Pred.No. 9,26-22; Matches 84; Conservative 30; Mismatches 90; Indels 10; Gaps 4;

```
OY      226 RAPELKEQVSVKVKAAEDLMGTDSDLMFGYYIQOSHWOIFGNKNRPFRVHDYOPEILTQ   285  
       | :|||::          : | : ||| :: ||| :: ||| :: ||| :: ||| ::  
Db      83 RKDEKFQLSLIAFLPMRIGLLGPSNVLGSASYOKSWMWLNSKESPPRETRYEPDLFGE    142  
  
OY     286 PV-YEDELMPDGVKRVRIGGAIVHHSNGESAKLRSRWNRAYLMAAGMEWKKLVMPRIWGRIFF  344  
       | : : : | : | : | : ||| : | : | : | : | : | : | : | : | : | :  
Db     143 ATDIYFAAW--TLRVVEKGNYHNDSNGSRDPTRTNRLITRLAMENGMWLYEKFMRYI-   199  
  
OY     345 KEGSGSQDDNDPDLIDYYGYGDVFELYOLENKSNISGTFRVNPRSGCALOLDVVYP LGK   404  
       ||||||| | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db     200 -----GSTDDNPDLIRKYGGYOLKIGYHL-GEAYLAKAGQNMNTVGCGAEVGLSPVTK   253  
  
OY     405 GISGFYLFPGYGOSLDINYNHATSFCYGLMLND 438  
       : | : : |||:||||| | |||:|||:  
Db     254 HVRITYQVISGESLIDYNFNQTVRGVGMALND 287
```

RESULT 3

PAL_ECOLI STANDARD; PRT: 269 AA.

AC P00631:

DJ 21-JUL-1986 (Rel. 01, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DR 20-AUG-2001 (Rel. 40, Last annotation update)

DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DEREGENT-RESISTANT

DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPATIDIYCHOLINE 1-

ACTIVHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).

CN PLDA OR B3821 OR Z5342 OR EC84751.

GN Escherichia coli, and

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteriota; gamma subdivision; Enterobacteriaceae;
CC Escherichia.

OX NCBI_TaxId=562, 83334;

RN [1]

RP SEQUENCE FROM N.A.
RC MEDLINE=83157492; Pubmed=6397464;
RA Homma H., Kobayashi T., Chiba N., Karasawa K., Mizushima H., Kudo I.,
RA Inoue K., Ikeda H., Sekiguchi M., Nojima S.;
RT "The DNA sequence encoding plda gene, the structural gene for
RT detergent-resistant phospholipase A of E. coli.";
RL J. Biochem. 96:1655-1664(1984).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX Medline=92358234; Pubmed=1379743;
RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT "Analysis of the Escherichia coli genome: DNA sequence of the region
RT from 84.5 to 86.5 minutes.";
RL Science 257:771-778(1992).
RN [3]
RP REVISION TO 14-15.
RP STRAIN-K12 / MG1655;
RX MEDLINE=97426617; Pubmed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perne N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.C., Rose D.O.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; Pubmed=11206551;
RA Perne N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postal G., Hackitt J., Klinsk S., Boulton A., Shao Y., Miller L.,
RA Grobeck E.U., Davis N.W., Lim A., Dimilanta E.T., Potamousis K.,
RA Apodaca E., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."

RL Nature 409:529-533(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-0157:H7 / RIMD 0509952;
 RA MEDLINE-01576331; PubMed-11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
 Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 Kuhara S., Shiba T., Hattori M., Shingawa H.,
 RA "Complete genome sequence of enterohemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [6]
 RP SEQUENCE OF 30-289 FROM N.A.
 RC STRAIN-K12;
 RA MEDLINE-85003590; PubMed-6383820;
 RA de Geus P., Verheij H.M., Riegman N.H., Hoekstra W.P.M., de Haas G.H.;
 RT "The pro- and mature forms of the E. coli K-12 outer membrane
 RT phospholipase A are identical.";
 RL EMBO J. 3:1799-1802(1984).
 RN [7]
 RP SEQUENCE OF 174-289 FROM N.A.
 RC STRAIN-K12;
 RA MEDLINE-87115164; PubMed-3027506;
 RA Irlino N., Nakayama K., Nakayama H.;
 RT "The rego gene of Escherichia coli K12: primary structure and
 RT evidence for SOS regulation.";
 RL Mol. Gen. genet. 205:298-304(1986).
 RN [8]
 RP MUTAGENESIS OF SER-172.
 RA MEDLINE-94131966; PubMed-8300539;
 RA Brok R.G.P.M., Brinkman E., van Bortel R., Bekkers A.C.A.P.,
 RA Verheij H.M., Tomassen J.;
 RT "Molecular characterization of enterobacterial pldA genes encoding
 RT outer membrane phospholipase A.";
 RL J. Bacteriol. 176:861-870(1994).
 RN [9]
 RP ACTIVE SITE SER-164, AND PARTIAL SEQUENCE.
 RA MEDLINE-91249806; PubMed-2040286;
 RA Horrevorts A.J.G., Verheij H.M., de Haas G.H.;
 RT "Inactivation of Escherichia coli outer-membrane phospholipase A by
 RT the affinity label hexadecanesulfonyl fluoride. Evidence for an
 RT active-site serine.";
 RL Eur. J. Biochem. 198:247-253(1991).
 CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
 CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
 CC REQUIRED FOR EFFICIENT SECRETION OF BACTERIOCINS, SEEMS TO BE
 CC DORMANT IN NORMAL GROWING CELLS.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
 CC LOCATED THERE.
 CC -1- INDUCTION: BY MEMBRANE DAMAGING, FOR EXAMPLE, BY PHAGE-INDUCED
 CC LYSIS OR TEMPERATURE SHOCK.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X02143; CAA26081.1; -
 CC EMBL: M87049; AAA67617.1; -
 CC EMBL: AE000458; AAC76824.1; -
 CC EMBL: AE005613; AAG59017.1; -
 CC EMBL: AP002567; BAB38174.1; -
 CC EMBL: M30198; AAA24516.1; -
 CC PIR: A00771; PSECA.

DR PIR: A22133; PSECA.
 DR PIR: S30711; S30711.
 DR Ecogene: EG10738; PldA.
 DR InterPro: IPR003187; PldA.
 DR Pfam: PF02253; PldA; 1.
 KW Hydrolyase; Lipid degradation; Outer membrane; Signal; Calcium;
 KW Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 289 PHOSPHOLIPASE A1.
 FT ACT_SITE 164 164
 FT MOTIF 172 172 S->F: INACTIVE PROTEIN.
 FT CONFLICT 14 15 LP -> FA (IN REF. 2).
 FT CONFLICT 30 33 DAPA -> MTRQ (IN REF. 6).
 SQ SEQUENCE 289 AA; 33163 MW; A688AD32AA60F218 CRC64;
 Query Match 15.1%; Score 357; DB 1; Length 289;
 Best Local Similarity 39.3%; Pred. No. 1.le-21;
 Matches 84; Conservative 29; Mismatches 91; Indels 10; Gaps 4;
 QY 226 RAPLAKQVSVKVAADLMQTSDFYTGQSHQQLFNKNSRRPRVHYOPEIFLQ 285
 DB 83 RKDEVKFKQLSLAPFLMGLGPNVSLGASTYQKSMQLSNSSESPRETNYEPQLGLF 142
 QY 286 PV-YSDLPMQGVKVMIGAVHNSGESAKLSRSMNRAYLAGEMKNLTVMPRMGRIF 344
 DB 143 ATDYRFAGW--TLNDVEMKYNHDSNGSDPTSRSMNRLYTLMAENGWLVKVPWYV- 199
 QY 345 KEGSGSQDDNPDLIDYGGYGVDFLYQLENKSNISGTVRYNPRSGKALQLDYVPLGK 404
 DB 200 -----GMTDDMPDITKMGYQLKIGYHL-CDAVLSAKGQYNNWTVYGGAGELGSLPYTK 253
 QY 405 GISGFQIFQCGOSGLDYHNEATSFVGLMLND 438
 DB 254 HVRLYTVGSGESGLIDYFNQTRVGVMLND 287
 RESULT 4
 ID PAL_PROVU STANDARD; PRT; 289 AA.
 AC P37447;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DEREGENT-RESISTANT
 DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
 DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
 GN PldA.
 OS Proteus vulgaris.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Proteus.
 OX NCBI_TaxID=585;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-94131966; PubMed-8300539;
 RA Brok R.G.P.M., Brinkman E., van Bortel R., Bekkers A.C.A.P.,
 RA Verheij H.M., Tomassen J.;
 RT "Molecular characterization of enterobacterial pldA genes encoding
 RT outer membrane phospholipase A.";
 RL J. Bacteriol. 176:861-870(1994).
 CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
 CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
 CC LOCATED THERE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X76902; CAA54224.1; -.
DR PIR: C36971; C36971.
DR PIR: S40130; S40130.
DR InterPro: IPR003187; PLA1.
DR Pfam: PF02253; PLA1. 1.
KW Hydrolase; Lipid degradation; Outer membrane; Signal; Calcium.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 289 PHOSPHOLIPASE A1.
FT ACT_SITE 164 164 BY SIMILARITY.
FT ACT_SITE 164 164 PHOSPHOLIPASE A1.
SQ SEQUENCE 289 AA; 32944 MW; D75516CFPB406997 CRC64;

Query Match 13.6%; Score 321.5; DB 1; Length 289;
Best Local Similarity 33.3%; Pred. No. 8.3e-19;
Matches 87; Conservative 42; Mismatches 115; Indels 17; Gaps 7;

QY 179 LDRNNPTLMSRPHNPVYLPIFMHGKPRSPNTPSHEARQTPNEFRAPELKFGVSKV 238
DB 43 LQEHNP-FTLVPRESNTLYTY-----TSDLNKKAIESYNSDNA-NKDEYKFLSLAF 95
QY 239 KAEDLMGTDSDLFMFGYTOQSHWQIFNGKNSRPFVHYDQPEIFLTPQV-YSDLPWDGKV 297
DB 96 PLMRGILGNSLILGASYTORSMWOLSTNGESAPFRETYEPOLFGLFATDYSVGWV--TL 153
QY 298 RMIGGAVHNSGESAKLSRSNBRAYLMAGMEKMLTVPRIRMGRIEFGKSSGSDPDND 357
DB 154 KRAEFGYHNSGSRSDPRTSRNNRLXSLMAONGMLVEKFWYI-----GDTSDMN 207
QY 358 ILDYGYGDVRLYOLENKSINISGTVRYNPRSGKALQLDVYVPLGKISGYFOJFGYG 417
DB 208 IRYKGYGYLKGIGYQI-GEAVLSAKGYVMWNTGSGAELGVSYPITKHVRYFTQYSGYG 266
QY 418 QSLIDYINHEATSFYGLMLND 438
DB 267 ESLIDYDFNQTRVGMGMVLD 287

RESULT 5
GUN_BACSI
ID GUN_BACSI STANDARD: PRT; 800 AA.
AC P06564;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (ALKALINE CELLULOSE).
OS Bacillus sp. (strain 1139).
OC Bacteria; Firmicutes; Bacillales; Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1411;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=87085443; Pubmed=3098909;
RA Fukumori F., Kudo T., Narahashi Y., Horikoshi K.;
RT "Molecular cloning and nucleotide sequence of the alkaline cellulase
RT gene from the alkalophilic Bacillus sp. strain 1139."
RL J. Gen. Microbiol. 132:2329-2335(1986).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- MISCELLANEOUS: ALKALOPHILIC BACILLUS SP., STRAIN 1139, PRODUCES
CC ONE ALKALINE CELLULOSE (PH OPTIMUM 9.0). THIS STRAIN IS NOT A TRUE
CC CELLULOTIC MICRO-ORGANISM BECAUSE THE ENZYME IS UNABLE TO
CC HYDROLYSE NATIVE CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC ..
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M15743; AAA22305.1; -.
DR EMBL: D00066; BAA00045.1; -.
DR PIR: A29003; A29003.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR Pfam: PF001050; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5_1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 800 ENDOGLUCANASE.
FT ACT_SITE 190 190 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 305 305 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 800 AA; 88602 MW; 7CCA4D7BBDAD55CF CRC64;

Query Match 4.6%; Score 109.5; DB 1; Length 800;
Best Local Similarity 19.8%; Pred. No. 0.47;
Matches 95; Conservative 54; Mismatches 163; Indels 169; Gaps 25;

QY 9 TSLIISCFAILAIQAKAPNVAFAVDYRSEND--LGDNELPIDVQASATQASASTDA 65
DB 7 TQLISSLILVL-LLSLFTALAEGRNTREDNFKHLGLNDVYKRSEAGALQLOEVDQ 65
QY 66 NPL-DEHEPELY-----TTALENKTMLINCSAL-----NODIMRLACY-DTLVHGEP 111
DB 66 MTLVQHGKIDLRKMSHGLQWFPFELINDNAYKALANDMESMIRLAMYVEENGYSANP 125
QY 112 AVIKTKRSIRIDEIT-----WQT-----IKGRQVYV 138
DB 126 ELIKSRVIGIDILAIENDMYVIVDWHVHAPGDRPDVYAGADEFRDLAALYPNPHIY 185
QY 139 QETDPR-----IFLMGKEGMLTKKAKOLEYAAKOTPLSLSPFDLDRN---NTPLW 187
DB 186 ELANEPSSNNNGAGCPNNEEGMAVAK-----EYADPIYEMLRDGNMADNIIIVGSPW 240
QY 188 SSRP-----HNPMVLPPIF--MHGKPNRS--PNTPSHEARQTPNEFRAPELKQ 233
DB 241 SQRPDLADNPIDDHHTYTVHFGYGSNAHSFESYRPETPNBERGNVMSNTYTA----LE 296
QY 234 VSVKVKAAEDLMGT-----DSDLMFGYTOQ-----SHWQIFNGKNS-----RP 271
DB 297 NCVAVFATE--WGTSQANGDGPRFDEADWIEFLNENNISWANSLSLV-KNEVSGAFRP 353
QY 272 FVHV-----DYOPEIFLTPQVYSDLPW---D 294
DB 354 FELGKSNATSLDPGPDQWVPEELSLSGEYVARIRKGVYEP--IDRKYYKVLMDEND 410
QY 295 GKVRMIGCAVHNSGES-----AKLSRSNBRAYLMAGMEKMLTVPRIRMG 341
DB 411 GTRKGQGV-----NGDSVEDVYIENEGALKLSGLDASNDVSGEYVANNRLSADNG 464
QY 342 R 342
DB 465 K 465

RESULT 6
AAMD_YEAST
ID AAMD_YEAST STANDARD: PRT; 810 AA.
AC P15274;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE AMP DEAMINASE (EC 3.5.4.6) (MYOADENYLATE DEAMINASE).
GN AMD1 OR AMD OR YML035C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
```

```

OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-90105403; PubMed-2690949;
RA Meyer S.L., Kvalnes-Rick K.L., Schramm V.L.;
RT "Characterization of AMD, the AMP deaminase gene in yeast. Production
of amd strain, cloning, nucleotide sequence, and properties of the
protein."
RL Biochemistry 28:8734-8743(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-9288C; AB972;
RA Badcock K., Churcher C., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: AMP DEAMINASE PLAYS A CRITICAL ROLE IN ENERGY
    METABOLISM.
CC -1- CATALYTIC ACTIVITY: AMP + H(2)O = IMP + NH(3).
CC -1- PATHWAY: PURINE NUCLEOTIDE CYCLE.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SIMILARITY: BELONGS TO THE ADENOSINE AND AMP DEAMINASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M30449; AAA34420.1; -
DR EMBL: Z46559; CAA6620.1; -
DR PIR: A33365; A33365.
DR SGD: S0004498; AMD1.
DR InterPro: IPR001365; A_deaminase.
DR Pfam: PF00962; A_deaminase; 1.
DR PROSITE: PS00485; A_DEAMINASE; 1.
DR KMW HydroLase: Nucleotide metabolism.
FT ACT_SITE 422 422 POTENTIAL.
FT ACT_SITE 631 631 POTENTIAL.
FT ACT_SITE 707 707 POTENTIAL.
FT ACT_SITE 708 708 POTENTIAL.
FT ACT_SITE 568 568 F -> C (IN REF. 1).
FT CONFLICT 568 568 F -> C (IN REF. 1).
SQ SEQUENCE 810 AA; 93301 MW; 7A6DCB43B9B45C93 CRC64;

```

```

Query Match 4.28; Score 98.5; DB 1; Length 810;
Best Local Similarity 21.38; Pred. No. 3.7;
Matches 74; Conservative 39; Mismatches 132; Indels 103; Gaps 15;

```

```

OY 46 DNELPIDVOSATOSASDTAN-----PLDEHEPELTTLAENTML 86
DB 48 DEHPLEQDSDSHESLADSNANFSTYENQOQLBENGKOLALDBHDS--HSAILEOPSHS 105
OY 87 INCSALNODIMRLACYDTLVH-----GETPAVITKTRSRILDETTIMQTIKGPQVYVQET 141
DB 106 TNCSSSIAMAMNG-HDSADHASNGSGKPRITLSASQHLIPETL-KSFAGAPVYNNQVR 163
OY 142 TDPIFLMG-----NEKGMILTKKAKOLEVAKOFTPLSLSFDDRRNTP 186
DB 164 TSASYKMGMLADDAOQFLDDPSSSELIDLYSKVAECNLRKAKYQTTISVQNDONPKKPKG 223
OY 187 W--SSRPHNPMY-----VLPIFMHGKPNRSPMPRSHEARQFTNPNRAELKQVSVKV 238
DB 224 WVVVPPRPKSTYSIDTKTVV-----TKNPKDAEVDFTKCE-----260
OY 239 KAAEDLMGTDSDLMFGYTOOSHQIFENGKNSRPFVADYQPEIFLTQPVYSDLPMDGKVR 298
DB 261 -----IPGEDPDMETLNDSDYVV-----HRSGKTDELLAQIPTLADYLDLE-----K 305
OY 299 MIGHGAVHNSNGESAK-----LSRSNRRAYILMAGMEKKNLTPMR 338
DB 306 MISIS-----SDGPAKSPAYRRLOYLEARNMLYYLL--NEYQETSVSRR 347

```

```

RESULT 7
ID BCK1_YEAST STANDARD; PRT; 1478 AA.
AC Q01389; P32894;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SERINE/THREONINE PROTEIN KINASE BCK1/SLK1/SSP31 (EC 2.7.-.-).
GN BCK1 OR SLK1 OR SSP31 OR LAS3 OR YJ095W OR J0906.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92186847; PubMed-1545797;
RA Costigan C., Gehring S., Snyder M.;
RT "A synthetic lethal screen identifies SLK1, a novel protein kinase
homolog implicated in yeast cell morphogenesis and cell growth."
RL Mol. Cell. Biol. 12:1162-1178(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-92104496; PubMed-1840547;
RA Irie K., Araki H., Oshima Y.;
RT "A new protein kinase, SSP31, modulating the SMP3 gene-product
involved in plasmid maintenance in Saccharomyces cerevisiae."
RL Gene 108:139-144(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-BG123;
RX MEDLINE-92107166; PubMed-1729597;
RA Lee K.S., Levin D.E.;
RT "Dominant mutations in a gene encoding a putative protein kinase
(BCK1) bypass the requirement for a Saccharomyces cerevisiae protein
kinase C homolog."
RL Mol. Cell. Biol. 12:172-182(1992).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN-S288C;
RX MEDLINE-95176706; PubMed-7871887;
RA Miosga T., Boles A., Schaaff-Gerstenschlaeger I., Schmitt S.,
RA Zimmermann F.K.;
RT "Sequence and function analysis of a 9.74 kb fragment of
Saccharomyces cerevisiae chromosome X including the BCK1 gene."
RL Yeast 10:1461-1488(1994).
RN [5]
RP SEQUENCE OF 602-1104 FROM N.A.
RA Cusick M.E.;
RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: SERINE/THREONINE PROTEIN KINASE INVOLVED IN A SIGNAL
    TRANSDUCTION PATHWAY THAT PLAY A ROLE IN YEAST CELL MORPHOGENESIS
    AND CELL GROWTH. THIS PATHWAY SEEMS TO STARTS BY SMP3; THEN
    INVOLVE THE KINASE PC1 THAT MAY ACT ON THIS KINASE. BCK1 PROBABLY
    PHOSPHORYLATES MKR1 AND MKR2 WHICH THEMSELVES PHOSPHORYLATE THE
    MPK1 KINASE.
CC -1- SUBCELLULAR LOCATION: CYTOSOL; MITOCH. (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MAP KINASE KINASE KINASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M84389; -; NOT_ANNOTATED_CDS.
DR EMBL: D10389; BAA01226.1; -
DR EMBL: X60227; CAA42788.1; -
DR EMBL: X77923; CAA54896.1; -
DR EMBL: Z49370; CAA89389.1; -

```

```

DR EMBL: Z49369; CAA89388.1; -.
DR EMBL: M88604; AAA21179.1; -.
DR PIR: S20117; S20117.
DR PIR: S22285; S22285.
DR PIR: J01118; J01118.
DR PIR: J01432; J01432.
DR HSSP: P24941; 1A01.
DR SGD: S0003631; BCK1.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002290; Ser_thr_kin_ataste.
DR Pfam: PF00069; pkinase; 1.
DR SMART: SMO0220; S_TKc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Transferrase: Serine/threonine-protein kinase; Tyrosine-protein kinase;
KW ATP-binding; Phosphorylation.
FT DOMAIN 1175 1440 PROTEIN KINASE.
FT NP_BIND 1181 1189 ATP (BY SIMILARITY).
FT BINDING 1204 1204 ATP (BY SIMILARITY).
FT ACT_SITE 1303 1303 BY SIMILARITY.
FT MOD_RES 1134 1134 PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT VARIANT 1119 1119 T -> P (IN BCK1-19; ACTIVATION).
FT VARIANT 1120 1120 I -> K (IN BCK1-11; ACTIVATION).
FT VARIANT 1120 1120 I -> T (IN BCK1-16; ACTIVATION).
FT VARIANT 1120 1120 G -> V (IN BCK1-20; ACTIVATION).
FT VARIANT 1146 1146 A -> P (IN BCK1-20; ACTIVATION).
FT VARIANT 1174 1174 A -> I (IN REF. 2).
FT CONFLICT 59 59 E -> I (IN REF. 3).
FT CONFLICT 79 79 A -> P (IN REF. 3).
FT CONFLICT 264 264 A -> P (IN REF. 3).
FT CONFLICT 279 279 N -> I (IN REF. 3).
FT CONFLICT 703 714 RYPTPTSYIDR -> STPKRVITME (IN REF. 3).
FT CONFLICT 795 795 S -> A (IN REF. 3).
FT CONFLICT 802 802 L -> V (IN REF. 3).
FT CONFLICT 808 808 A -> S (IN REF. 3).
FT CONFLICT 903 903 T -> N (IN REF. 3).
FT CONFLICT 919 919 T -> N (IN REF. 3).
FT CONFLICT 960 962 ADA -> RDR (IN REF. 5).
FT CONFLICT 1086 1104 RPVPDSSEYEFIDGLNCK -> VPIAHTSSYRMDLVKIN
H (IN REF. 5).
SQ SEQUENCE 1478 AA; 164194 MW; D586C3A497A5BB3 CRC64;

```

```

Query Match 4.1%; Score 97.5; DB 1; Length 1478;
Best Local Similarity 23.0%; Pred. No. 10;
Matches 110; Conservative 50; Mismatches 175; Indels 143; Gaps 26;

```

```

OY 2 KVSITLFLS-----ILSCFALIAIOAKAVNPVAFVDEVSENDLQDNELPI---DV 53
DB 912 KVNNSNSTVSTNSLTFYSPSPILKRGNSKRYVSSISAAD-IFEENDITFADAPPFDSDD 970
OY 54 OSANOSASTDTANPLDEHEPELYTTALENK--TYMLINCSALNODIMRLACYDTLVHGETP 111
DB 971 SDDSSSSDDDIIMSKKKAPE---TNNENKKDEKSDNSTSHDELF-----QTO 1019
OY 112 AVIKTKBSIRLDEITWQITKKGKPOVYVQETDPIFLMGNKGMILTKKAKOLEVYAAKQFT 171
DB 1020 DKMERK-----MFRPSPEVYVO-NLEKFFRANLTKPIF-----EGIASPTS 1061
OY 172 PLTISFDLDRNNT-----PLWSSRPHPMYVLPFIMHG-----KPNSPNTP----- 213
DB 1062 PKSLDLSLSPKNVASSKREPESTPSRPVPPDSSYEFTIQGLNGKNKPLQAKTPKTKTIR 1121
OY 214 --SHEARQFTNEFRAPRLKFOVSQVKAFAE-DLWGTSDMLFQGTQOSHWOIFGNKNSR 270
DB 1122 TIAHEASLARKN-----SVYKLKRMQTKMGT--RMVEYTELHNHNSINKAKKSK 1167
OY 271 PFRVHDYPELFLQPVYSDLPMDQKVMIGKA-----VHNSNGSAKLSR----- 317
DB 1168 -----GEKKEFAW-MKGEMIKGSGAVYLLANVTGEMAAVKQVEVPKY 1211
OY 318 -SWNRAYLMAGEMKKNLVMPRIMGRIRKEKSGSGSDPNRPILDVYGGVDRFLYLQLENK 376

```

```

DB 1212 SQONFALLS-----TV-----EALSENSTLKDLHLNIVQYLG-----ENK 1249
OY 377 SNISGTVRYNPRSGKALQDLYVPLGKIGSGYFOIFQYGSGLIDYVHEATSPFGVL 434
DB 1250 NNIV-----SLFLEYV--AGSGVSLIMYGRFDEPLI--KHLTQYVLKGL 1291

RESULT 8
ID PBPB_STRPN STANDARD; PRT; 719 AA.
AC 004707;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE PENICILLIN-BINDING PROTEIN 1A (PBP-1A) (EXPORTED PROTEIN 2).
GN POMA OR EXP2.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=45607; AND 63915;
RX MEDLINE=93010977; PubMed=1396576;
RA Martin C., Sibold C., Hakenbeck R.;
RT "Relaxedness of penicillin-binding protein 1a genes from different
RT clones of penicillin-resistant Streptococcus pneumoniae isolated in
RT South Africa and Spain."
RL EMBO J. 11:3831-3836(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92325042; PubMed=1624444;
RA Martin C., Briese T., Hakenbeck R.;
RT "Nucleotide sequences of genes encoding penicillin-binding proteins
RT from Streptococcus pneumoniae and streptococcus oralis with high
RT homology to Escherichia coli penicillin-binding proteins 1a and 1b."
RL J. Bacteriol. 174:4517-4523(1992).
RN [3]
RP SEQUENCE OF 293-369 FROM N.A.
RC STRAIN=RMX;
RX MEDLINE=95020625; PubMed=7934910;
RA Pearce B.J., Yin Y.B., Maure H.R.;
RT "Genetic identification of exported proteins in Streptococcus
RT pneumoniae."
RL Mol. Microbiol. 9:1037-1050(1993).
CC -1- FUNCTION: CELL WALL FORMATION.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- MISCELLANEOUS: THESE STRAINS ARE PENICILLIN-SENSITIVE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X67873; CAA48073.1; -.
DR EMBL: X67872; CAA48072.1; -.
DR EMBL: M90527; AAA26956.1; -.
DR PIR: S28038; S28038.
DR InterPro: IPR001264; Transglycosyl.
DR InterPro: IPR001460; Transpeptidase.
DR Pfam: PF00912; Transglycosyl; 1.
DR Pfam: PF00905; Transpeptidase; 1.
DR PRODOM: PD001895; Transglycosyl; 1.
KW Peptidoglycan synthesis; Antibiotic resistance; Cell wall;
KW Multifunctional enzyme.
FT ACT_SITE 370 370 ACYLATED BY PENICILLIN (BY SIMILARITY).
FT DOMAIN 657 683 SER-RICH.
FT VARIANT 124 124 T -> A (IN STRAIN R6).
FT VARIANT 386 386 V -> I (IN STRAIN 63915).

```

FT VARIANT 388 388 D -> E (IN STRAIN R6).
 FT VARIANT 397 397 E -> K (IN STRAIN 63915).
 FT VARIANT 523 523 M -> I (IN STRAIN 63915).
 FT VARIANT 533 533 E -> D (IN STRAINS 63915 AND R6).
 FT VARIANT 540 540 T -> S (IN STRAINS 63915 AND R6).
 FT VARIANT 657 657 S -> N (IN STRAIN 63915).
 SQ SEQUENCE 719 AA: 79745 MW: 58C36A93BFA970A CRC64;

Query Match 4.0%; Score 95; DB 1; Length 719;
 Best Local Similarity 20.1%; Pred. No. 6;
 Matches 96; Conservative 70; Mismatches 196; Indels 116; Gaps 25;

QY 3 VLSLTLLSI-----LSCFALLIAQKAVRPVAVDRENDL 43
 DB 13 LSTLSLTVIAALVILGGVFYVSKAPSLSEKLVATTSSKYNKNDLIADGSERRY 72
 QY 44 -GDNDLPIDVOSATOS-----ASTDTANPLDEHEBELYTTALENTKMLINCSALN 93
 DB 73 NQANDIPIDLVKAIYSIEDHREFDRGIDTIRILGAFILNLSNLSGG-----STLT 126
 QY 94 ODIMRLACDTLVHGETPRAVIKTKRSIRLDETITWOTIKGPOVVOET---TDPIFLMG 149
 DB 127 OOLIKLTFTSTSDQTI-----RKAQEAFLAIQLEOKATKOEILTYINKYMSN 178
 QY 150 NEKGMLTK-----KAKOLE-----YAAKOTPLS--LSFDDLRNTPIMSSRP 191
 DB 179 GNVMQTAANQNYKQDNLNLSLPOLALLAGMPQAPNQYDPSHPREAQDRNLVL--SEM 236
 QY 192 HNPVYLPIFMHGKPNRSPNTPSHEARQFTPNFRAPE-----LKFQVS--VKAKAEDLW 245
 DB 237 KNGYI-----SAEQYKAVNPTITDGLQSLKASNPAYMDNLKEVINVEETGYNL 292
 QY 246 GTDSDLMEGTQ--QSH--NOIENGKNSRPRVHDYQPEIFLQPVYSDLPWDGKVRMIGM 302
 DB 293 TFGMDVYTVNDQAKHLDIYNTDEVYAVPDELO-----VASTIVDS--NGKV--IAQL 345
 QY 303 GAVVHNSN-----GESAKLSRSNNRAYLWAGMKMLT--VMPRI-----WRIKREGS 348
 DB 346 GARHSSNNSFGINQAVETNRDM-----GSTMKPITDYAPALEGYVDSTATYIDEP 398
 QY 349 GSGQPDNDPILDY-YGY-GDVRELYOLENKSNSIGTVRYN-----PRSGKALQLDY 398
 DB 399 YNPGTPTFYVMNDRGYFGNITLQYVALQSSRNVPVETLNKVKLNKAKFLNGLGIDY 456

RESULT 9

GUN_PAPEO STANDARD: PRT: 397 AA.
 ID GUN_PAPEO P23548:
 AC 01-NOV-1991 (Rel. 20, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ENDOGUCANASE (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE).
 OS Paenibacillus polymyxa (Bacillus polymyxa).
 OC Bacteria: Firmicutes: Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Paenibacillus.
 OK NCBI_TaxID=1406;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90170877; PubMed=2307659;
 RA Baird S.D., Johnson D.A., Seligy V.L.;
 RT "Molecular cloning, expression, and characterization of
 endo-beta-1,4-glucanase genes from Bacillus polymyxa and Bacillus
 circulans";
 RL J. Bacteriol. 172:1576-1586(1990).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: M33791; AAA2631.1; -
 DR PIR: A35136; A35136.
 DR HSSP: P54583; 1ECE.
 DR InterPro: IPR001547; Glyco_hydro_F5.
 DR Pfam: PF00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation: Hydrolase; Glycosidase.
 FT ACT_SITE 194 194 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 317 317 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 397 AA: 44357 MW: B9C2E802C04F0A2A CRC64;

Query Match 4.0%; Score 94.5; DB 1; Length 397;
 Best Local Similarity 25.7%; Pred. No. 2.9;
 Matches 46; Conservative 21; Mismatches 61; Indels 51; Gaps 10;

QY 301 GMGAVHNSNGESAKLSRSNNRAYLWAGMKMLTVMPRIWGR-----IFREG----- 347
 DB 43 GNKIVDESGKEAFLNGLNW-----FGLETPNTY-LHGLMSRSMDDMLDQVKKEGYNLIR 95
 QY 348 -----SGSQPD-----NDIILDYGYGDVRELYOLENKSNSIGTV-----RINPRS 389
 DB 96 LPSYNQLFDSSSRPDSIDYKKNPDLV---GLNPDIQIMDKLEKAGRGIOIILDRHRPGS 152
 QY 390 GKCALQLDYV--YPLGKISGYFOIFQGY-----GQSLIDVNHETSRGVLMDNW 439
 DB 153 G-GQSELWITSQYIPESRWISDMKMLADRYKNPNPTVIGADLHNPFGQASMGTONASTDW 210

RESULT 10

BAC2_MOUSE
 ID BAC2_MOUSE P97303;
 AC 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TRANSCRIPTION REGULATOR PROTEIN BACH2 (BTF AND CNC HOMOLOG 2).
 GN BACH2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C;
 RX MEDLINE=97042438; PubMed=8887638;
 RA Oyake T., Itoh K., Motohashi H., Hayashi N., Hoshino H., Nishizawa M.,
 RA Yamamoto M., Igarashi K.;
 RT "Bach proteins belong to a novel family of BTF-basic leucine zipper
 RT transcription factors that interact with Max and regulate
 RT transcription through the NF-E2 site";
 RL Mol. Cell. Biol. 16:6083-6095(1996).
 CC -1- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT ACTS AS REPRESSOR OR
 CC ACTIVATOR. BINDS, IN-VITRO, TO NF-E2 BINDING SITES. PLAY IMPORTANT
 CC ROLES IN COORDINATING TRANSCRIPTION ACTIVATION AND REPRESSION BY
 CC MAX.
 CC -1- SUBUNIT: HETERODIMER OF BACH2 AND MAX.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: EXPRESSION RESTRICTED TO MONOCYTES AND
 CC NEURONAL CELLS.
 CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. CNC SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 1 BTF/POZ DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D86604; BAA1338.1; -
 DR HSSP: P05412; IFOS.
 DR MGD: MGI:894679; Bacch2.
 DR InterPro: IPR000210; BTB_POZ.
 DR InterPro: IPR001871; bZIP.
 DR Pfam: PF00651; BTB; 1.
 DR Pfam: PF00170; bZIP; 1.
 DR SMART: SM00338; BRLZ; 1.
 DR SMART: SM00225; BTB; 1.
 DR PROSITE: PS50097; BTB; 1.
 DR PROSITE: PS00036; bZIP_BASIC; 1.
 KW Transcription regulation; Activator; Repressor; DNA-binding;
 KW Nuclear protein.
 FT DOMAIN 37 103 BTB.
 FT DNA_BIND 162 168 POLY-GLU.
 FT DOMAIN 527 542 BASIC MOTIF.
 FT DOMAIN 550 572 LEUCINE-ZIPPER.
 SQ SEQUENCE 716 AA; 78935 MW; 913283731AE24333 CRC64;

Query Match 4.0%; Score 94; DB 1; Length 716;
 Best Local Similarity 23.1%; Pred. No. 7.2; Mismatches 140; Indels 156; Gaps 24;
 Matches 101; Conservative 40;

OY 14 SCFAIIAIQAKAVNPVAFV---DEV---RSENDG-----QDNELPIDVQSATQSAST 62
 DB 124 SCSFSLQQLNLNREG--LFVCRKDSACORQEDHNSAGEEHEEETMDSTAMACAT 181
 OY 63 D-----TANPLDEH-----EPELYTALNKTMLNCSALNO---DIMRLAC 101
 DB 182 DQMLDPDISFEATAPVAKEKALLPESEVPTDTKENSEK---GALIQYPRKKYQLAC 237
 OY 102 ---YDILVHG-----ETPAVTKTKRSTL---DET 125
 DB 238 TKNVSAHSHTGSGFASFSDESPGNSLKLPGCLPMQIKSEPSSEETESTLTLCSGET 297
 OY 126 IWQITKGRPOVYQETTPPI-----FLMGNEKMLTKK 158
 DB 298 ---DIKDRPGDVEMRKQPSAPRTSTRTGAACIDRSRSVSSPCLRLSLFGITKGV--- 350
 OY 159 DAKOLEYAKQFTPLSLF-----DIDRNNPLMWS--RPHNPMVYLPIFMHGK 205
 DB 351 ESTGIPSTYQQ--PLVRSACPFNKGISQGLKTDYTLAGNYGQPHYGOKDVSNFAMGS 408
 OY 206 PNRSP-----NTPSHEARQFTPNFRAPELK-----FQVSVKVAEAEDLWGTDSL 251
 DB 409 PLRGGPETLCFFSSPCSGARFLATEHQBGLMGDGMGNVNOVRPOIK-CEOSYGTNSSD 467
 OY 252 WGYTQASHMOIFNGKNSRPFVRVHDYQPEIFLTQPV--YSDLPMGKRYMIGMGAVHHSN 309
 DB 468 ESG-----SEFSADSESCVQDROGEVKLPFVQDITDPLRNDPQMIMK---HKLIT 516
 OY 310 GESAKL-----SRSMNR 321
 DB 517 SEQLFEFIDIRRRSKNR 533

RESULT 11
 DP3A_BACSU STANDARD; PRT; 1115 AA.
 AC 034623;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DNA POLYMERASE III ALPHA SUBUNIT (EC 2.7.7.7).
 GN DNAE.
 OS Bacillus subtilis.
 OS Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/staphylococcus group; Bacillus.

OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE:9804867; PubMed:9387221;
 RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
 RT "Sequencing and functional annotation of the Bacillus subtilis genes
 in the 200 kb rnb-dnaB region";
 RL Microbiology 143:3431-3441(1997).
 CC -1 FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
 CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
 CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
 CC THE ALPHA CHAIN IS THE DNA POLYMERASE (BY SIMILARITY).
 CC -1 CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
 CC N PYROPHOSPHATE + DNA(N).
 CC -1 SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
 CC CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
 CC DIMERIZATION TO FORM THE POLIIT' COMPLEX. POLIIT' ASSOCIATES WITH
 CC THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
 CC AND CHI) AND WITH THE BETA CHAIN (BY SIMILARITY).
 CC -1 SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1 SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-C FAMILY. DNAA
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF008220; AAC00338.1; -
 DR EMBL: Z99118; CAB14883.1; -
 DR Subtilist; BG12583; dnaE.
 DR InterPro: IPR003141; PHP.N.
 DR InterPro: IPR002309; tRNA-synt_2.
 DR Pfam: PF02231; PHP.N; 1.
 DR Pfam: PF01336; tRNA_ant1; 1.
 DR SMART: SM00481; POLIITAC; 1.
 KW Transferase; DNA-directed DNA polymerase; DNA replication;
 KW Complete proteome.
 SQ SEQUENCE 1115 AA; 125349 MW; E8B04E339E512FE CRC64;

Query Match 4.0%; Score 94; DB 1; Length 1115;
 Best Local Similarity 21.7%; Pred. No. 13;
 Matches 100; Conservative 62; Mismatches 155; Indels 144; Gaps 26;

OY 20 AIOQKAYPN-PVAFVDEVREND-----LGQDNELPIDVQSATQSASTDTAN- 66
 DB 437 ADQLAKLIPSRGWTLDARQOSPOLDRLRESSILQGVYSTARKTEBLPRHASTHAGV 496
 OY 67 -----PLDEHEBELYTT--ALENKTMLNCSALNODIMRLACYDPLVHGSETPAV 113
 DB 497 VLSEBPLDVVPLQGHGCIYLTQYAMDH---LEDLGLLKMDFGLRL-PLISEST-SM 551
 OY 114 IKTKRSIRLDETIWQITKGRPOVYQETTPFLMGNEKMLTKKDAK--OLEYAA--- 167
 DB 552 IEKENIKIDLS-----SISYSD--DKTF-----SLLSKGDYTGIGOLESGAGMRS 594
 OY 168 --KQFTPLSLPDDRNNTPLMSSRPHNPMVYLPIFMHGKPNRSP-NTPSHEARQFTPNE 224
 DB 595 VLKRLKPSGLE-DIVAANA---LYRP-GPMENITPLEIDRKHGRAVHYPHEDLKRILEDT 649
 OY 225 F-----RAPELKPOVSVKVAEEDLWGTDSLWFGYTGQASHMQ 262
 DB 650 YGVITYGQIMMISRMAGFSLGEADLLRAVSKKKRIID-----RRSHF- 696
 OY 263 IFNGKNSRPFVRVHDYQPEIFLTQPVYSDLPMGKRYMIGMGAVHHSNGESAKLSRSNRA 322
 DB 697 -VEGLAKKEYGV-DTANENV-----DLIVKPFANYG-----FNRSHAVA 732
 OY 323 YLMAGMEKKNLTVMRIMGRIFKESGSGQPDNDPILDIYGV---GDVRFLYQLENNSN 378

```

DB 733 YSNIQCO---LAVLKAHYPLFCMLTISVIGNEKDISQVLYAKSGIRLPPVSKSS 789
QY 379 I-----SGTVRYNPRSGKALQLDIVYPLGKISGYFOIQ 414
DB 790 FPETVENGSVRYSLRAIKSV-----GVSAAVKDIYK 819

RESULT 12
SRK6_BRAOL STANDARD; PRT; 849 AA.
ID SRK6_BRAOL STANDARD; PRT; 849 AA.
AC 009092;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PUTATIVE SERINE/THREONINE KINASE RECEPTOR PRECURSOR (EC 2.7.1.37)
DE (S-RECEPTOR KINASE) (SRK).
GN SRK6.
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucotyledons; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3712;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. S656; TISSUE=Stigma;
RC MEDLINE=92020942; PubMed=1681543;
RA Stein J.C., Howlett B., Boyes D.C., Nasrallah M.E.;
RT "Molecular cloning of a putative receptor protein kinase gene encoded
RT at the self-incompatibility locus of Brassica oleracea."
RL Proc. Natl. Acad. Sci. U.S.A. 88:8816-8820(1991).
CC -1- FUNCTION: INVOLVED IN SPOROPHYTE SELF-INCOMPATIBILITY SYSTEM
CC (THE INABILITY OF FLOWERING PLANTS TO ACHIEVE SELF-
CC FERTILIZATION), PROBABLY ACTING IN COMBINATION WITH S-LOCUS-
CC SPECIFIC GLYCOPROTEINS. INTERACTION WITH A LIGAND IN THE
CC EXTRACELLULAR DOMAIN TRIGGERS THE PROTEIN KINASE ACTIVITY OF THE
CC CYTOPLASMIC DOMAIN.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN -> ADP + A PHOSPHOPROTEIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY IN THE PISTIL AND ANTHER.
CC -1- POLYMORPHISM: THERE ARE A NUMBER OF DIFFERENT S ALLELES IN
CC B. OLERACEA, POSSIBLY PROVIDING THE RECOGNITION SPECIFICITY.
CC -1- SIMILARITY: THE EXTRACELLULAR DOMAIN IS SIMILAR TO S-LOCUS
CC GLYCOPROTEINS OF BRASSICA, WHILE THE INTRACELLULAR DOMAIN IS
CC A SER/THR-PROTEIN KINASE RELATED TO RAF KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: M76647; AAA33000.1; AL7_TERM.
DB HSSP: P1362; IFGI.
DR InterPro: IPR001460; B_lectin.
DR InterPro: IPR00719; Euk_pkinase.
DR InterPro: IPR003609; Pan_app.
DR InterPro: IPR002290; Ser_thr_kin_actsite.
DR InterPro: IPR000858; Slocus_glycop.
DR Pfam: PF000954; Slocus_glycop; 1.
DR SMART: SM00108; B_lectin; 1.
DR SMART: SM00473; Pan_AP; 1.
DR SMART: SM00221; SYRK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Transferrase: Serine/threonine-protein kinase; Signal; ATP-binding;
KW Transmembrane; Receptor; Glycoprotein; Self-incompatibility.
FT SIGNAL 1 32
FT CHAIN 33 849 PUTATIVE SERINE/THREONINE KINASE

```

```

FT DOMAIN 33 446 RECEPTOR.
FT TRANSMEM 447 466 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 467 467 POTENTIAL.
FT DOMAIN 528 779 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 534 542 ATP (BY SIMILARITY).
FT BINDING 556 556 ATP (BY SIMILARITY).
FT ACT_SITE 653 653 BY SIMILARITY.
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 389 389 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 442 442 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 849 AA; 97231 MW; 7E156059EDFA370 CRC64;

Query Match 3.9%; Score 92.5; DB 1; Length 849;
Best Local Similarity 21.9%; Pred. No. 12;
Matches 78; Conservative 50; Mismatches 107; Indels 121; Gaps 24;

QY 105 LVHGETPAVIRKTRSRIRLDETTIQTIGKPOVYVYQETTDPIFLMG---NEK---GMLTK 157
DB 24 LIHPALSIYINT-----LSSTESLFTISSNKTLY---SPGISFVGFPRTRSRWYLGMYK 75
QY 158 K-DAKOLEYAKOPTPLSLSD-----LDRNNTPLMSRRPNNPYVLPPIFHHCK 205
DB 76 KVSADRYVVAANDPNPLSNIGTLKISGNLVLDHSNKNKVMWNTN-----LTRGN 125
QY 206 PNSSP-----NTPSEARQPT-----PNEFRABELKFOVSKYKAAEDLW 245
DB 126 -EESPVAELLANGNFMVRRSSNNNDASEYIMOSEPDPTDLLEMLGLYMLKT----- 177
QY 246 GTDSDLMEFYTQSHMOIFNGKNSRPFVHDYQ-----PEILFTQ---PVYSDLPWDGK 296
DB 178 GLNRFL-----TSWRSSDPSSGNF---SYKLETSLEPEFLYSRENFPMHRSRGPWG- 226
QY 297 VRHIGKCAVH-----HSGESAKLSRSNNRAV-----LMAGEMKNLVMF--RI 339
DB 227 IRRSGIPEDOKLSYMYNFLENNEEVAAYTFPMNNSFSLTLTISEYFQRLTWYPSIRI 286
QY 340 WGRIFEGSGSDPPDNDIIDL-----YGYGVRFYLTLENKSNISGTV--VYPRS 389
DB 287 WNRFW-----SSPYD-PQCDTYIMCGPYATCDY-----NTSPVCNCIOGFNPRN 329

RESULT 13
MET_MOUSE
ID MET_MOUSE STANDARD; PRT; 1379 AA.
AC P16056; O62125;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HEPATOCYTE GROWTH FACTOR RECEPTOR PRECURSOR (EC 2.7.1.112) (MET PROTO-
DE ONCOGENE TYROSINE KINASE) (C-MET) (HGF RECEPTOR) (HGF-SF RECEPTOR).
GN MET.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=88262253; PubMed=2838789;
RA Chan A.M.L., King H.W.S., Deakin E.A., Tempest P.R., Hilkens J.,
RA Kroezen V., Edwards D.R., Wills A.J., Brookes P., Cooper C.S.;
RT "Characterization of the mouse met proto-oncogene."
RL Oncogene 2:593-599(1988).
RN [2]
RP SEQUENCE OF 1199-1270 FROM N.A.
RC MEDLINE=90152381; PubMed=2482828;
RA Wilks A.F., Kurban R.R., Hovens C.M., Ralph S.J.;
RT "The application of the polymerase chain reaction to cloning members

```


RT of the protein tyrosine kinase family.";
 RL Gene 85:67-74(1989).
 CC -1- FUNCTION: RECEPTOR FOR HEPATOCYTE GROWTH FACTOR. HAS A TYROSINE-
 CC PROTEIN KINASE ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP + PROTEIN
 CC TYROSINE PHOSPHATE.
 CC -1- SUBUNIT: HETERODIMER FORMED OF AN ALPHA CHAIN (50 KDA) AND A BETA
 CC CHAIN (145 KDA) WHICH ARE DISULFIDE LINKED.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- DISEASE: ACTIVATION OF MET AFTER REARRANGEMENT WITH THE TPR
 CC (TRANSLOCATED PROMOTER) LOCUS OF CHROMOSOME 1 PRODUCES AN
 CC ONCOGENIC PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE MET TYROSINE KINASE FAMILY OF
 CC RECEPTORS.
 CC -1- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y00671; CAA68680.1; -;
 DR EMBL: M33424; AAA40015.1; -;
 DR PIR: S01254; S01254.
 DR HSSP: P11362; 1f61.
 DR MGD: MGI:96969; Met.
 DR InterPro: IPR000719; Euk_PKinase.
 DR InterPro: IPR002909; IPT_TIG.
 DR InterPro: IPR003659; PSI.
 DR InterPro: IPR002165; Plexin_repeat.
 DR InterPro: IPR001627; Sema.
 DR InterPro: IPR001245; Tyr_Kin.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF01437; Plexin_repeat; 1.
 DR Pfam: PF01403; Sema; 1.
 DR Pfam: PF01833; TIG; 3.
 DR PRINTS: PR00109; TYRKINASE.
 DR SMART: SM00429; IPT; 4.
 DR SMART: SM00423; PSI; 1.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transferrase; Tyrosine-protein kinase; Proto-oncogene; ATP-binding;
 KW Receptor; Transmembrane; Glycoprotein; Phosphorylation; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 1379 HEPATOCYTE GROWTH FACTOR RECEPTOR.
 FT DOMAIN 25 931 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 932 954 POTENTIAL.
 FT DOMAIN 955 1379 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 55 499 SEMA.
 FT DOMAIN 1076 1343 PROTEIN KINASE.
 FT SITE 306 307 CLEAVAGE (POTENTIAL).
 FT NP_BIND 1082 1090 ATP (BY SIMILARITY).
 FT BINDING 1108 1108 ATP (BY SIMILARITY).
 FT ACT_SITE 1202 1202 BY SIMILARITY.
 FT MOD_RES 1233 1233 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 357 357 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 606 606 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 634 634 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 784 784 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 878 878 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1199 1199 V -> I (IN REF. 2).
 FT CONFLICT 1199 1255 V -> R (IN REF. 2).
 FT CONFLICT 1251 1261 K -> T (IN REF. 2).
 FT CONFLICT 1261

FT CONFLICT 1269 1270 VL -> IP (IN REF. 2).
 SQ SEQUENCE 1379 AA; 153548 MW; FC5CC87FDD8ADE8 CRC64;
 Query Match 3.9%; Score 92.5; DB 1; Length 1379;
 Best Local Similarity 21.3%; Pred. No. 24;
 Matches 80; Conservative 48; Mismatches 123; Indels 125; Gaps 22;
 QY 39 SENDAGDNELEPT-DVQASQTASASTDTANPDHEHRELYTTALENKTMLINCALSALNODIM 97
 DB 1014 SQNCACRQVQPLDPLDLSLISGSDISPL-----LQN-TYHIDLSALNPELV 1061
 QY 98 RLACDTLVHGETPAVITKTSIRLDETIWOTIKGKPOVVOETDPIFLMGNEKMLTK 157
 DB 1062 QAVQH-VVIGPSSSLI-----VHEVEVIG-RGHFCVYV-----GTLILD 1097
 QY 158 KDAKOLEAAKQFPLSLSPDLDRNPTPLMS-----SRPP-----NPWYV 197
 DB 1098 NDGKRIHCAVNSLNRIIT---DIEEVSQFLTEGIIMKDSHPNVLSLGLCLRSBSPLVY 1154
 QY 198 LPIFMHGKPNMSPNTPSHEARQFPNEFRABELKQVSVKKAADLWGTDSLMFGYTO 257
 DB 1155 LPYMKHG-----DLRNFIRETHNPYVK-----DLIG-----FGL- 1184
 QY 258 QSHWQIFNGKN---SRPRVADYQPEIFLTQPVYSDLFPMDGKVRNIGM-----AV 305
 DB 1185 ---QVARGMKYLSAKKFEVHRD---LAARCMLEDEKFTKVVADFGIARDMYDEKYSY 1235
 QY 306 HHSNGESAKLSRWNRRAVLMGMEWKNLTVMPRMG-----RIPEKSGSQPDNDI 358
 DB 1236 HNKTS-AKLEPVK---WALSLSLOQKFTTSDVMSFVGLMELMTRGAPRPDPVNTDI 1290
 QY 359 LDYGYGCVREPLYOLE 374
 DB 1291 TTYLLQG--RRLLQPE 1304
 RESULT 14
 YL70_ARCFU STANDARD; PRT; 324 AA.
 ID YL70_ARCFU
 AC 028112;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN AF2170.
 GN AF2170.
 OS Archaeoglobus fulgidus.
 CC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 CC Archaeoglobus.
 CC NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Kienk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kierlavage A.R., Graham D.E., Kyriplids N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
 RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RA "The complete genome sequence of the hyperthermophilic, sulphate-
 RA reducing archaeon Archaeoglobus fulgidus.";
 RT Nature 390:364-370(1997).
 RL Nature 390:364-370(1997).
 CC -1- SIMILARITY: CONTAINS 2 KELCH REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AEO00955; AAB89093.1; -
DR TIGR: AE2170; -
DR InterPro: IPR001798; Kelch.
DR Pfam: PF01344; Kelch; 2.
KW Hypothetical protein; Repeat; Complete proteome.
FT REPEAT 229
FT REPEAT 276
FT REPEAT 323
SQ SEQUENCE 324 AA; 36025 MW; C445388CFEB96E45 CRC64;

Query Match 3.9%; Score 92; DB 1; Length 324;
Best Local Similarity 20.1%; Pred. No. 3.5;
Matches 54; Conservative 44; Mismatches 72; Indels 98; Gaps 18;

OY 224 EFAPPELKFQVSVKVAED-----LMGTSDLMEGYTQSHWOLF-NGKNSRPVRH 275
DB 63 EFPPN-RLEI---LSSSEFQREAAVWDGELLIFGT-----VFENGKTSPTDQIL 112
OY 276 DYQPEIFLTPYVSDLP-----W-DGKYRMI-----GKAVHNSNGESAKLSRS 318
DB 113 SFNPKLERLKLVLNLSLPHPTSDVAAVWGDGRVYIFLNNSRCEYAYAPNESPAKLDS 172
OY 319 -----W-NRAYL-----MAGHEKMLTVMPRIWGRIRKESGSGQ 351
DB 173 CPLEHGGCVSVYVWGKAYFECGEVASFDPMGFKM--IAFTDVRRAATVADG-- 228
OY 352 PDNDPDLIDYGVDRFLYLENKSINSGT-----VRYNRSKGALQDLVYVPLGK-- 405
DB 229 -----YIAIGSSCIAETKDEILIRFNKTBG-LOEMTKLPVARGQA 270
OY 406 -ISG-YFOIF-----QGYGOSLIDYNNH 425
DB 271 VAVGGEYIYFPGYTKDGYANETLRDYD 298

RESULT 15
YJHO_YEAST
ID YJHO_YEAST STANDARD; PRT; 888 AA.
AC P40361;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHEICAL 104.3 KDA PROTEIN IN SMC3-MRPL8 INTERGENIC REGION.
GN YJL070C OR J1095 OR HRD888.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycas.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RA STRAIN=5288C;
RA Sor F.J.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: BELONGS TO THE ADENOSINE AND AMP DEAMINASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z34288; CAA84052.1; -
DR EMBL: Z49345; CAA89362.1; -
DR EMBL: Z49344; CAA89361.1; -
DR EMBL: X88851; CAA61309.1; -
DR PIR: S47120; S47120.
DR SGD: S0003606; YJL070C.
DR InterPro: IPR001365; A.deaminase.
DR Pfam: PF00962; A.deaminase; 1.
KW Hypothetical protein; Hydrolase.
SQ SEQUENCE 888 AA; 104263 MW; DB31A8086224114D CRC64;

Query Match 3.9%; Score 92; DB 1; Length 888;
Best Local Similarity 22.1%; Pred. No. 14;
Matches 66; Conservative 36; Mismatches 126; Indels 70; Gaps 15;

OY 128 QTIKGRQVYVQE-----TTDPIFLMGNEKGLTKKDA-----KOLEYAKOFTPLSL 176
DB 2 QAVERRPSLLFDEQNSVTAKPNETKNKEARVLSNDGDVSPVLKQKEISVDDMTSLP 61
OY 177 FDLDRN---NTPLW---SSRPHNPMYVLPPIFMHGRPNR-----SPNTPSH-----EAROF 220
DB 62 TEPDRQVAVLSPMFEDELEENKIDPLPSVSHHNGNESDSFVSSTPPSNLKTGEETKDL 121
OY 221 TPNEFRAPPELKFQVSVKVAEDLWG-----TSDLMEGYTQ--QSHWOLFNGKNSRPR 273
DB 122 FINPF---ELVSQMRKRYIAASKODGISNINKNDEKMFVLPKPLPKWRFEDDK-----R 173
OY 274 VHDYQPEIFLTPYVSDLPMDGKVRMIGKAVHNSNGESAKLSKMRATYLMAGMEKML 333
DB 174 FQD-----PSDSDLNDGDSTGTGAATPRHNGYVPSYFTDHYYYTTRKSLGKGN 224
OY 334 TVMPRIWGRIRKESGSGSDPDNDPDLIDYGVDRFLYLENKSINSGTFRVYRNPRSGK 391
DB 225 IKVP-YTGEYF-----DLEDY-----KKQYTHLSNONTNPL--SPYSSK 263

Search completed: November 30, 2001, 14:27:05
Job time: 552 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:26:27 ; Search time 77.9 Seconds

(without alignments)
829,941 Million cell updates/sec

Title: US-09-787-083-6

Perfect score: 2360

Sequence: 1 MKVSLSTLTLSLSCFALLA.....YNHEATSPFGVGLMDNMGL 442

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP mhc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	815	34.5	382	2 09K007	09K007 neisseria m
2	815	34.5	409	2 09Jr21	09Jr21 neisseria m
3	373.5	15.8	306	2 09Ct22	09Ct22 pasteurella
4	366	15.5	329	2 09PM08	09PM08 campylobact
5	360.5	15.3	292	2 09S1D7	09S1D7 yersinia ps
6	360	15.3	289	2 09L6N9	09L6N9 salmonella
7	356.5	15.1	292	2 09Z4N8	09Z4N8 enterobacte
8	342.5	14.5	297	2 032349	032349 campylobact
9	246.5	10.4	355	2 025241	025241 helicobacte
10	240.5	10.2	355	2 09ZLX5	09ZLX5 helicobacte
11	123	5.2	278	2 09XB53	09XB53 erwinia car
12	114	4.8	1686	4 000443	000443 homo sapien
13	110.5	4.7	602	11 063485	063485 rattus norv
14	109.5	4.6	821	2 059241	059241 bacillus sp
15	108	4.6	824	2 09F216	09F216 bacillus sp
16	105.5	4.5	901	5 018749	018749 caenorhabdi
17	104	4.4	783	2 045554	045554 bacillus sp
18	104	4.4	798	4 09UN32	09un32 homo sapien
19	104	4.4	798	4 09UBK2	09ubk2 homo sapien

20	103.5	4.4	3247	12 065553	065553 bovine herp
21	103	4.4	435	2 09X9C0	09X9C0 streptococc
22	101.5	4.3	403	5 P91736	P91736 hydra magni
23	101.5	4.3	1046	2 084941	084941 streptococc
24	101	4.3	4307	5 019319	019319 caenorhabdi
25	100.5	4.3	1509	11 061194	061194 mus muscucu
26	100	4.2	719	2 057114	057114 streptococc
27	100	4.2	719	2 09REr8	09REr8 streptococc
28	99.5	4.2	739	5 09V9B6	09V9B6 drosophila
29	99.5	4.2	1658	11 061182	061182 mus muscucu
30	99	4.2	719	2 054948	054948 streptococc
31	99	4.2	719	2 054946	054946 streptococc
32	99	4.2	719	2 09WVW0	09WVW0 streptococc
33	99	4.2	719	2 09WV11	09WV11 streptococc
34	99	4.2	719	2 09REr6	09REr6 streptococc
35	99	4.2	719	2 09REr4	09REr4 streptococc
36	99	4.2	727	2 09F2G4	09F2G4 streptococc
37	99	4.2	1372	10 09FL92	09FL92 arabidopsis
38	99	4.2	5005	2 09PZ5	09PZ5 ureaplasma
39	98.5	4.2	467	4 09UJMS	09UJMS homo sapien
40	98.5	4.2	467	4 09BYT2	09BYT2 homo sapien
41	98.5	4.2	576	2 09L115	09L115 streptomyc
42	98.5	4.2	660	2 09KGR8	09KGR8 bacillus ha
43	98.5	4.2	788	10 09SCV5	09SCV5 arabidopsis
44	98	4.2	719	2 054947	054947 streptococc
45	98	4.2	719	2 054949	054949 streptococc

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	382 AA.
09K007	09K007			
ID	09K007			
AC	09K007			
DT	01-OCT-2000 (TREMblrel. 15, Created)			
DT	01-OCT-2000 (TREMblrel. 15, Last sequence update)			
DT	01-JUN-2001 (TREMblrel. 17, Last annotation update)			
DE	PHOSPHOLIPASE A1, PUTATIVE.			
GN	NMB0464.			
OS	Neisseria meningitidis (serogroup B).			
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.			
OX	NCBI_TaxID=491;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MC58 / SEROGROUP B;			
RX	MEDLINE=2015755; PubMed=10710307;			
RA	Tettein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,			
RA	Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,			
RA	Nelson W.C., Gwyn M.L., DeBoy R., Peterson J.D., Hickey E.K.,			
RA	Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,			
RA	Mason T., Ciecko A., Parksey D.S., Blair E., Ciltone H., Clark E.B.,			
RA	Cotton M.D., Ulteback T.R., Khouri H., Qin H., Yamahavean J.,			
RA	Gill J., Scarlato V., Maignani V., Pizzo M., Grandi G., Sun L.,			
RA	Smith H.O., Fraser C.M., Moxon E.R., Rappunli R., Venter J.C.;			
RT	"Complete genome sequence of Neisseria meningitidis serogroup B strain			
RT	MC58."			
RL	Science 287:1809-1815(2000).			
DR	EMBL: AE002403; AAF40901.1; -.			
DR	TIGR: NMB0464; -.			
DR	InterPro: IPR003187; PLAI.			
DR	Pfam: PF02253; PLAI: 1.			
KW	Complete proteome.			
SO	SEQUENCE 382 AA; 42714 MW; B468A802F062E836 CRC64;			
Query Match	34.5%;	Score 815;	DB 2;	Length 382;
Best Local Similarity	44.7%;	Pred. No. 1.6e-62;		
Matches 163;	Conservative 67;	Mismatches 111;	Indels 24;	Gaps 8;
QY	87	INCSALNODIRKLACVDTLVGERTPAVI-----KTKRSIRLDETIWQTI-KGKPOVYVQOE 140		
DB	33	LGCAALFDNVRILACVDIRFAAQLPSSAGGEGESKAVLNTLTVRSSLDGGEAVIYVEK 92		


```

DE PHOSPHOLIPASE A (EC 3.1.1.32).
GN PLDA OR C11351.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIC 11168;
RC MEDLINE=20150912; PubMed=10689204;
RA Parhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagsis K., Kariyasev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Randleman M.A., Rutherford K.M., Van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
DR EMBL: AL139078; CAB7378.1;
DR InterPro: IPR003187; Pfam:
DR Pfam: PF02253; Pfam: 1.
DR PRINTS: PR01486; PHPHLIPASEA1.
KW Complete Proteome.
SQ SEQUENCE 329 AA; 38880 MW; A32731F2B751AC44 CRC64;

Query Match 15.5%; Score 366; DB 2; Length 329;
Best Local Similarity 33.2%; Pred. No. 1,2e-23;
Matches 105; Conservative 48; Mismatches 125; Indels 38; Gaps 13;

OY 130 IKGKPOVVOETTPDPIFLMGNEKGM.LTKKDAQLEVAAGQFPLSLSPDIDRNT--PLW 187
DB 43 LKNSSVLISQEQONNSQATQTQNSTITKEKO-----DFSLALANTIGENESFNPLG 96
OY 188 SSRPHNPVPLPIFMHGKPNRSPNPSHEAROPTNEFRAPELKFOVSYVKAEDLMGT 247
DB 97 IS-SYKMYFLP-FAYSFNSLGVNNKSEA-----KFLQSVYKRLFFENLGL 141
OY 248 DSDLMFGYQQOSIMQIFNGKNSRPFVHDYQPEIFLTQPVY-SDLPMWDGKVRNIGAGVH 306
DB 142 DEKYLAAYVQTSWQIV--EHSSPFRETNYQPFETIDPLYLKDYEFENLR--VGILH 196
OY 307 HSNGESAK--LSKSNRAYLMAEMKNTLVMPRIKRIKESGSGQDDNDPILDYGY 364
DB 197 ESNKGDEMLQSSWNRIVYFATILNKLFLVRLMVRT--PENKKDDNPAILHMGV 253
OY 365 GGVRFYQLQENKSNISGTVRYNPR--SGKALQLDYVPL-GKISGYQIFOGYQSGLI 421
DB 254 FQVNLAY-LGDDYFIMLRLNNLKFNNKGAIGVDIGDIFNGGIWYLQYFNGESLI 312
OY 422 DYNHEATSFQVGLMLN 437
DB 313 DYNKHLQRLSTGFLIS 328

RESULT 5
O9SID7 PRELIMINARY: PRT: 292 AA.
AC O9SID7;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PHOSPHOLIPASE A.
GN PLDA.
OS Versinia pseudotuberculosis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Versinia.
OX NCBI_TaxID=633;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PIII PIBI;
RC Kariyasev A.V., Winzeler E.A., Williams K.J., Oyston P.C.,
RA Tildall R.W., Wren B.W.;

```

```

RT "Biochip-based Signature-Tagged Mutagenesis: Identification and
RT characterisation of Y. pseudotuberculosis gene plda essential for
RT virulence in mice."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ245393; CAB51586.1;
DR InterPro: IPR003187; Pfam:
DR Pfam: PF02253; Pfam: 1.
DR PRINTS: PR01486; PHPHLIPASEA1.
SQ SEQUENCE 292 AA; 33758 MW; 8E712D908ACB5A5 CRC64;

Query Match 15.3%; Score 360.5; DB 2; Length 292;
Best Local Similarity 39.2%; Pred. No. 3e-23;
Matches 83; Conservative 34; Mismatches 86; Indels 9; Gaps 5;

OY 229 ELKFQVSVKKAEDLMGTDSDLMFGYQTSQSHQIFNGKNSRPFVHDYQPEIFLTQPV- 287
DB 86 EVKFOLSLAPPIWRGIGDSSLGASVYQSRSMQASNSSESSPFRETNYEPQLFLAWSTD 145
OY 288 YSDLPMDGKVRMIGMGAVHHSNGESAKLSRSNRAYLMAEMKNTLVMPRIKRIKES 347
DB 146 YELAGW--TFREVEFEGFNHOSNGKADPTSRSMNRYTVRYVAQKGNLEIDLPWIRPESD 203
OY 348 SSGSQDDNDPILDYGYGQVDFLYQL-ENKSNISGTVRYNPRSGKALQLDYVPLGKI 406
DB 204 S---KDDNPDTIKMYGKYLKVGATLQDSVPSLDG--RYWMNMGYGAEMGMSYPIRKHV 258
OY 407 SGYFOIFQYGGSLIDYNHEATSFQVGLMLND 438
DB 259 RFTYQVFSYGESMIDYNFQTRVGVGLMLND 290

RESULT 6
O9L6N9 PRELIMINARY: PRT: 289 AA.
AC O9L6N9;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PLDA PROTEIN.
GN PLDA.
OS Salmonella typhimurium LT2.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=99287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SGSC1412;
RA WashU;
RT "The Salmonella typhimurium Genome Sequencing Project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SGSC1412;
RA Waterston R.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF233324; AAF33435.1;
DR InterPro: IPR003187; Pfam:
DR Pfam: PF02253; Pfam: 1.
SQ SEQUENCE 289 AA; 32967 MW; DA97F5E1651C49C6 CRC64;

Query Match 15.3%; Score 360; DB 2; Length 289;
Best Local Similarity 39.7%; Pred. No. 3,3e-23;
Matches 85; Conservative 29; Mismatches 90; Indels 10; Gaps 4;

OY 226 RAELEKFOVSVKKAEDLMGTDSDLMFGYQTSQSHQIFNGKNSRPFVHDYQPEIFLTQ 285
DB 83 RKEVYKFOLSLAPPIWRGIGDSSLGASVYQSRSMQASNSSESSPFRETNYEPQLFLG 142
OY 286 PV-YSDLPMDGKVRMIGMGAVHHSNGESAKLSRSNRAYLMAEMKNTLVMPRIKRI 344
DB 143 ATDYRFAGW--TLRDVEMGYNHDSNGRSDPTSRSMNRYTVRYVAQKGNLEIDLPWIR 199

```


RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.; Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RT "The complete genome sequence of the gastric pathogen *Helicobacter*
 pylori.";
 RT Nature 388:539-547(1997).
 DR EMBL: AE000564; AAD07564.1; -
 DR TIGR: HP0499; -
 DR InterPro: IPR003187; PLAI.
 DR Pfam: PF02253; PLAI; 1.
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 355 AA; 42486 MW; 461960F15E65AB0A CRC64;

Query Match 10.4%; Score 246.5; DB 2; Length 355;
 Best Local Similarity 25.1%; Pred. No. 3, 2e-13;
 Matches 82; Conservative 41; Mismatches 117; Indels 87; Gaps 11;

OY 157 KKDQOLEYAAKQFTPLSFDDLRRNTPLMSSRPHNPVLPFIMHGKPNRSPNPSHE 216
 DB 69 KKLNMMDVIGTYFLPRYHSF-----TFIFQYHNPINP----- 102
 OY 217 AROFTPNEFRAPLKFQVSVKAAEDLWGTSDLMFGYQOSHWOIFNGKNSRPFVRVD 276
 DB 103 ---YORNEF-----KFOISFRVPEVFRHILMTKGTLYLAQTDMFOYNDPOSAPMNMN 154
 OY 277 YOREIFLTQPVYSDLPWDGV---RMIGMAVHNSG--ESAKLSRSNRAYLMAHMKWN 332
 DB 155 FMEPLIYVYPI-NKPRFGKIGNFSIWMQHISNGVGAOCYOPNPK---EGNPENQ 209
 OY 333 LVTWPRP-----MGRIFEGSGSOP-----DDNP 357
 DB 210 FPGQPIVADYNGQKDYRWGCGRSVSAQRPVRLWKEGKLIMVAYWYVYDQSNPN 269
 OY 358 ILDYGYGVDFLY-----OLENKSNSGTVRYNPRSGKALQLDYVYPLKGISGYF 410
 DB 270 LIDYMGVYNAKIDYRRGRHNFELQLDYDFQYWRD--RWHGAFRLGYTYRINPFVGIYA 327
 OY 411 QIFQGYGOSLIDYNHATSFGVGLMLN 437
 DB 328 QWENGYDGLYEYDFVSNRIGVIRLN 354

RESULT 10
 O9ZLX5 PRELIMINARY; PRT: 355 AA.
 AC O9ZLX5;
 DT 01-MAY-1999 (Tremblrel, 10, Created)
 DT 01-MAY-1999 (Tremblrel, 10, Last sequence update)
 DT 01-JUN-2001 (Tremblrel, 17, Last annotation update)
 DE PUTATIVE PHOSPHOLIPASE A1.
 GN PLDA OR JHP0451.
 OS *Helicobacter pylori* j99 (Campylobacter *pylori* j99).
 OC Bacteria; Proteobacteria; epsilon subdivision; *Helicobacter* group;
 OC *Helicobacter*.
 OX NCBI_Taxid=85963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Molt D.T., King B.L., Brown E.D., Dolg P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyts G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen *Helicobacter pylori*.";
 RL Nature 397:176-180(1999).
 DR EMBL: AE001479; AAD06029.1; -
 DR InterPro: IPR003187; PLAI.
 DR Pfam: PF02253; PLAI; 1.
 KW Complete proteome.
 SO SEQUENCE 355 AA; 42329 MW; B3CE9810EBA2FDC9 CRC64;

Query Match 10.2%; Score 240.5; DB 2; Length 355;
 Best Local Similarity 25.1%; Pred. No. 1e-12;
 Matches 83; Conservative 44; Mismatches 109; Indels 95; Gaps 13;

OY 157 KKDQOLEYAAKQFTPLSFDDLRRNTPLMSSRPHNPVLPFIMHGKPNRSPNPSHE 216
 DB 69 KKLNMMDVIGTYFLPRYHSF-----TFIFQYHNPINP----- 102
 OY 217 AROFTPNEFRAPLKFQVSVKAAEDLWGTSDLMFGYQOSHWOIFNGKNSRPFVRVD 276
 DB 103 ---YORNEF-----KFOISFRVPEVFRHILMTKGTLYLAQTDMFOYNDPOSAPMNMN 154
 OY 277 YOREIFLTQPVYSDLPWDGV---RMIGMAVHNSG--ESAKLSRSNR----- 321
 DB 155 FMEPLIYVYPI-NKPRFGKIGNFSIWMQHISNGVGAOCYOPNPKPNPNOFPQ 213
 OY 322 -----ATLMAG-----MWNK--NLVTWPRMGRIFEGSGSQPDD- 354
 DB 214 PVIYKQVNGQKDYRWGCGRSVSAQNALCFVLMWKEGKLIMVAYWYV-----PYDQ 265
 OY 355 -NPDIIDYGYGVDFLY-----OLENKSNSGTVRYNPRSGKALQLDYVYPLKGIT 406
 DB 266 SNPQLIDYMGVYNAKIDYRRGRHNFELQLDYDFQYWRD--RWHGAFRLGYTYRINPFV 323
 OY 407 SGYFOIFQGYGOSLIDYNHATSFGVGLMLN 437
 DB 324 GIYAQWENGYDGLYEYDFVSNRIGVIRLN 354

RESULT 11
 O9XB53 PRELIMINARY; PRT: 278 AA.
 AC O9XB53;
 DT 01-NOV-1999 (Tremblrel, 12, Created)
 DT 01-NOV-1999 (Tremblrel, 12, Last sequence update)
 DT 01-NOV-1999 (Tremblrel, 12, Last annotation update)
 DE KDUT.
 GN KDUT.
 OS *Erwinia carotovora*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Pectobacterium*.
 OX NCBI_Taxid=554;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=ATCC 39048; GS101.
 RX MEDLINE=98065591; PubMed=9402024;
 RA McGowan S.J., Sebahia M., O'Leary S., Hardie K.R., Williams P.,
 RA Stewart G.S., Bycroft B.W., Salmood G.P.;
 RT "Analysis of the carbenem gene cluster of *Erwinia carotovora*:
 RT definition of the antibiotic biosynthetic genes and evidence for a
 RT novel beta-lactam resistance mechanism.";
 RT Mol. Microbiol. 26:545-556(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 39048; GS101.
 RX MEDLINE=98276484; PubMed=9614345;
 RA McGowan S.J., Bycroft B.W., Salmood G.P.;
 RT "Bacterial production of carbapenems and clavams: evolution of beta-
 RT lactam antibiotic pathways.";
 RT Trends Microbiol. 6:203-208(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 39048; GS101;
 RA McGowan S.J.;
 RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U17224; AAD38237.1; -
 SO SEQUENCE 278 AA; 31341 MW; 63769F4F3550E1B7 CRC64;

Query Match 5.2%; Score 123; DB 2; Length 278;
 Best Local Similarity 21.6%; Pred. No. 0.012;
 Matches 74; Conservative 52; Mismatches 128; Indels 88; Gaps 16;


```

0Y 29 NPAVAVDVRSNDGQONELPIDVQATOSASTDNPANDEHBEPEYTTALEKMTLIN 88
Db 87 NOVAVLTKKKELEAODRNIGT-----QOFTRAKEELEAEKRLIRT---NERLSOE 133
0Y 89 CSALNODIMRLACYDTLVHGETPAVITKTRSRIRDETIMO---TIKGRPOVYOE----- 140
Db 138 VEYLTEDVKRL--NEKTESNT---TKGELQTKIDELQASDVYKREKRLGCKELIHN 192
0Y 141 -----TDPIFLMGNEG-----MLTKKDAQOLEYAAKQFTPLSLSFDD 180
Db 193 ONSWINTELKTKTTELLALGREGKEWELTELCTENKEEDALISHESASPLS--SSP 251
0Y 181 RNNTPLMRSSRPHNPYVULPIFMHGKPNRSPTPSHEARQFTPNEFBARDELFGYQSVYKA 240
Db 252 NNLSPGTGMSQKTP-----VPRAQREAPSSGQENKKIRPROGRSSYWELEASEVM 300
0Y 241 AEDLMGTSDLMFGYTOOSH-----QIENGKNSRPRVHDYPER----- 281
Db 305 LSTRIGSGS---FGVUYGKGMHGDVAVILKLVVDPTEPOLQAFNEVAVLTKTRHVIIL 361
0Y 282 -----FLPO-----PVYSDL--PMDSKVMI-----GMGAVHNSGESA 313
Db 362 FMGYTKTKNLALIVTOMCGSSLYKHLHVOETKRFQMFOLIDAROTAOGMVLYLHAKNLIHR 422
0Y 314 KLSRSMNBAVYLMAGMEK-----NLVMPRIWGRIFKEGSGSQ----- 351
Db 422 DMKS--NNIFLHEGLTVIGDFGLATVKSRY-----SGSQVEOPTGSVLMAPAEYIR 472
0Y 352 -PDDNPDLL--DYVCYGDVRFYLOLEKNSNSGIVRVNPRSGKALODUYVPLGK 405
Db 473 MODNPNPSFOSDVSYGIV--LYEL-----MTGELPFSHINNROI-----IFWGRG 518

```

RESULT	14		
059241			
ID	059241	PRELIMINARY:	PRT: 821 AA.
AC	059241:		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)	
DE	ENO-1.4-BETA-GLUCANASE (EC 3.2.1.4) (CELLULASE) (ENDOGALACTANASE)		
DE	(CARBOXYMETHYL CELLULASE).		
OS	Bacillus sp.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;		
OC	Bacillus/Staphylococcus group; Bacillus.		
OX	NCHI_TaxID=1409;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-KSM-64;		
RA	Sunltomo N., Ozaki K., Ito S.;		
RL	Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.		
CC	-1 CATALYTIC ACTIVITY: ENDOMYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.		
CC	EMBL: M84963; AA073189.1; -.		
DR	HSSP; M85465; IA3H.		
DR	InterPro: IPR001547; Glyco_hydro_F5.		
DR	Pfam: PF00150; cellulase.1.		
KW	PROSITE: PS00659: GLYCOSYL__HYDROL_F5; UNKNOWN_1.		
KW	Hydrolase; Glycosidase		
SO	SEQUENCE 821 AA; 90910 MW; 73D438FEF0B40B5C CRC64;		

Query Match	4.68;	Score 109.5;	DB 2;	Length 821;
Best Local Similarity	19.88;	Pred. No. 0.81;		
Matches	95;	Conservative	54;	Mismatches 163;
				Indels 169;
				Gaps 25

```

Oy 9 TLSIISCAIILAIQAAKAVPMPVAFVDEVREND---LGOENELPIVQSTQASIDTA 65
    | : : : | | : | | | | | | | | | | | | | | | | | | | | | | |
Db 6 TKQISLSLILVL-LSLFPTALAAEGNTRDNRKHLGLGNVYKRPSEAGALQIQEYDQG 64
    | : : : | | : | | | | | | | | | | | | | | | | | | | | | | |
Oy 66 NPL-DEHEPELY-----TTALENTMLINCALS-----NQDMLRLACY-DLIVHGEP 111
    | : : : | | : | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db      65  MTLVNDGHEKIQGLQGMSTHGLQWPELLINDMAYKALANDMESNMIRLAMYVGENGYSANP 124
OY      112  AVIKTKRSIRIDETI-----WQT-----IKGPOVYI 138
           : : : : :
Db      125  ELIKSRVIKGIDILAIENDMVIYIMVHNAHPDPRDPYAGAEDEFPRDIALYPNPHIY 184
OY      139  QETIDP-----IFLKGNEGMLTKKDAKOLEYAKOFTPLISFPLDN-----NTPLY 187
           : : : : :
Db      185  ELANPESSNNNGAGIPNNEGMAVYK-----EYADPIVELRDSGNADNIIIVGSPMW 239
OY      188  SSRP-----HNPIYVLPiP--MHGKPNRS--FNTPSHAKOTTPMEFARPELKQ 233
           : : : : :
Db      240  SORPDLADNPIIDHHTWYVHVFYTGSHASTESYPEPITNSEKGNVMSNTRYA-----LE 295
OY      234  VSVKKKAEDIMGT-----DSDDLFGYTOQ-----SHWOJFNCKNS-----RP 271
           : : : : :
Db      296  NGVAVFATE--WGTSGQANGDGPPYFDEADWIEFLJENNISMAMWSLTTN--KNEVSGAFTP 352
OY      272  FRVH-----DROPELFFLTQPYVSDLPW-----D 294
           : : : : :
Db      353  FELGKSNAATSLDPCGPDQVWPEELISLGEYVRABIKVNEP-----IDRTKYVLMFEND 409
OY      295  GKVRMIGGAVHSHNGES-----AKLSRSMNRAYLTMAGMEKMLTVMPIRWG 341
           : : : : :
Db      410  GTKQGEFV-----NGDSPEVEDVYIENEGALKLTLSGLDASNDVSEGVYMWANARLSADGWG 463
OY      342  R 342
           : : :
Db      464  K 464

```

ID	Q9F216	PRELIMINARY;	PRT;	824 AA.
AC	Q9F216;			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	CELLULOSE.			
OS	Bacillus sp.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OX	Bacillus/Staphylococcus group; Bacillus.			
RN	NCBI_TaxID=1409;			
RP	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=KSM-S237;			
RX	MEDLINE=101036886; PubMed=11193393;			
RA	Hakamada Y., Hatada Y., Koike K., Yoshimatsu T., Kawai S.,			
RA	Kobayashi T., Ito S.;			
RT	"deduced amino acid sequence and possible catalytic residues of a			
RT	thermostable, alkaline cellulase from an alkaliphilic Bacillus			
RT	strain."			
RL	Biosci. Biotechnol. Biochem. 64:2281-2289(2000).			
DR	EMBL; ABO18420; BAB1360.1; "			
DR	InterPro; IPR001547; Glyco_hydro_F5.			
DR	Pfam; PF00150; cellulase.1.			
DR	PROSITE, PS00659; GLYCOSTL_HYDROL_F5; UNKNOWN.1.			
SO	SEQUENCE 824 AA: 91364 MW: 65FA940FE1D29E9 CRC64;			

Query Match	4.68;	Score 108;	DB 2;	Length 824;
Best Local Similarity	19.88;	Pred. No. 1.1;		
Matches	95;	Conservative	60;	Mismatches 159;
				Indels 166;
				Gaps 28;

```

OY      9 TLTLSITCAIILAQAKVAPVPVAFVEBVSREND---LGGDNEPLPIDVGSATSOASADTA 65
          | : : : |
Db       7 TKQLSSLLTILVL-LLSIFPALAAEGNTRRDNKHLGNDNVKRSEAGALQLEVDQG 65
          | : : : |
OY      66 NPL-DEHEPELY-----TTALENTMLINCALS-----NODIMRLACIDTLVHGE--- 109
          | : : : |
Db       66 MTLVNDGHEKLTOLKRMSTGHGLQFPEILLINDVAYKALSNMDWSNMIRLAMY---VEGNGY 121
          | : : : |
OY     110 --TPAVIKTK-----RSIRLDETI--WDN-----IKGRP 134

```

```
Db 122 ATNPelikQRIYDGIETLAIENDMYIVDMVHAPGDPDPDYAGAKDFPREIAALYPNNP 181
QY 135 QVVOETTD-----IFLMGNEKMLTKKDAKOLEYAAKQFTPLSFDLDRN-----N 183
Db 182 HIIVELANEPSSNNNGAGIPNNEGKAVK-----ETADPIVEMLRKSGNADNIIYVG 236
QY 184 TPLMSSRP-----HNPWYVLPITFMHGKPNRSPNTPSHANQFTPNFEFRAPELK-- 231
Db 237 SPNMSQRPDLADNPIDDHHTWYVHFYTGSHAASTESYPS-----TPNSERGNVMSNT 291
QY 232 ---FQSVKVKAAEDLMGT-----DSDLMPGYTQO-----SHWOIFNGKNS-- 269
Db 292 RYALENGVAVFATE--WGTSQASGDGPFEDADAVWIEFLNENNISWANWSLTN--KNEYS 348
QY 270 ---RPFRV-----HDYQP-EIFLP-----QPV-----YSDLPW-- 293
Db 349 GAFTEFELGKSNATNLDPGPDHVMAPPEELSLSGEYVRARIKGVNTEPIDRTKYTKVLMDF 408
QY 294 -DGKVRMIGMCAVHHNSGESAKLSRSNRAVYIMAGME-----WKNLTVMPRIWGR 342
Db 409 NDGTKQ--GFGVNSDSPNKEIAYDNENNNTLKVSGLDVSNVSDGNFWANARLSANGWGK 466
```

Search completed: November 30, 2001, 14:26:28
Job time: 565 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:26:25 ; Search time 77.9 Seconds
(without alignments)
829.941 Million cell updates/sec

Title: US-09-787-083-4
Perfect score: 2360
Sequence: 1 MKVSLSTLTSLTLPCEFLA.....YNHATSPGVGLMLNDWML 442

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP mhc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	815	34.5	382	2	09K007
2	815	34.5	409	2	09JTM21
3	373.5	15.8	306	2	09CL22
4	366	15.3	329	2	09PM08
5	360.5	15.3	292	2	09SID7
6	360	15.3	289	2	09LGN9
7	356.5	15.1	292	2	09ZAN8
8	342.5	14.5	297	2	032349
9	246.5	10.4	355	2	025241
10	240.5	10.2	355	2	09ZLX5
11	121	5.1	278	2	09XB53
12	109.5	4.6	824	2	09F216
13	109	4.6	821	2	059241
14	106.5	4.5	602	11	063485
15	106	4.5	798	4	09UM32
16	106	4.5	798	4	09UBK2
17	102.5	4.3	783	4	045554
18	102.5	4.3	1686	4	000443
19	99.5	4.2	1509	11	061194

20	98.5	4.2	467	4	09JMW5	09jmw5 homo sapien
21	98.5	4.2	467	4	09BRV2	09brv2 homo sapien
22	98.5	4.2	576	2	09L115	09l115 streptomyce
23	98.5	4.2	788	10	09SCV5	09scv5 arabidopsis
24	98.5	4.2	1658	11	061182	061182 mus musculu
25	98	4.2	435	2	09X9C0	09x9c0 streptococc
26	98	4.2	871	4	09Y5C3	09y5c3 homo sapien
27	98	4.2	938	4	09Y5F7	09y5f7 homo sapien
28	97.5	4.1	797	2	09RDM6	09rdm6 lactobacill
29	97	4.1	761	12	09WT88	09wt88 lt virus. o
30	96	4.1	422	2	09RC26	09rc26 streptomyce
31	95.5	4.0	389	2	P74515	P74515 synechocyst
32	95	4.0	523	4	09H856	09h856 homo sapien
33	95	4.0	739	5	09V9E6	09v9e6 drosophila
34	94.5	4.0	476	5	09B160	09b160 caenorhabdi
35	94.5	4.0	530	5	045879	045879 caenorhabdi
36	94.5	4.0	683	2	09A607	09a607 caulobacter
37	94.5	4.0	749	2	059154	059154 anaerocellu
38	94.5	4.0	772	2	0923V2	0923v2 pseudomonas
39	94.5	4.0	1046	2	084941	084941 streptococc
40	94.5	4.0	1641	2	09PDX7	09pdx7 xylella fas
41	94	4.0	454	11	09EG54	09eg54 mus musculu
42	94	4.0	455	4	012875	012875 homo sapien
43	94	4.0	765	2	054183	054183 streptomyce
44	93.5	4.0	564	3	094727	094727 schistosach
45	93.5	4.0	664	3	09C122	09c122 piromyces s

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	382 AA.
09K007	AC	09K007		
DT	01-OCT-2000 (TREMblrel. 15, Created)			
DT	01-OCT-2000 (TREMblrel. 15, Last sequence update)			
DT	01-JUN-2001 (TREMblrel. 17, Last annotation update)			
DE	PHOSPHOLIPASE A1, PUTATIVE.			
GN	NMB0464.			
OS	Neisseria meningitidis (serogroup B).			
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.			
OX	NCBI_Taxid=491;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MC58 / SEROGROUP B;			
RX	MEDLINE=20175755; PubMed=10710307;			
RA	Tetteijn H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,			
RA	Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,			
RA	Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,			
RA	Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,			
RA	Mason T., Ciecko A., Parksey D.S., Blair E., Ciltone H., Clark E.B.,			
RA	Cotton M.D., Utepbach T.R., Khouri H., Qin H., Yamathayan J.,			
RA	Gill J., Scarlato V., Maignani V., Piazza M., Grandi G., Sun L.,			
RA	Smith H.O., Fraser C.M., Moxon E.R., Rappunli R., Venter J.C.;			
RT	"Complete genome sequence of Neisseria meningitidis serogroup B strain			
RT	MC58."			
RL	Science 287:1809-1815(2000).			
DR	EMBL: AE002403; AAF40901.1; -.			
DR	TIGR: NMB0464; -.			
DR	InterPro: IPR003187; Pfam: PF02253; Pfam: 1.			
KW	Complete proteome.			
SQ	SEQUENCE 382 AA; 42714 MW; B468A802F062E836 CRC64;			
Query Match	34.5%;	Score 815;	DB 2;	Length 382;
Best Local Similarity	44.7%;	Pred. No. 6	Se-63;	
Matches 163;	Conservative 67;	Mismatches 111;	Indels 24;	Gaps 8;
QY	87	INCSALNODIMRLACVDTLVGERTPAVT----	KTKRSIRLDETITWQTI-KGKQVYVQOE	140
DB	33	LQCAALDNDVTRLACVDRIFAAQLPSSAGCGGSKAVLNTETVRSLSLDGEAVIVVEK	92	

Qy	141	TTDILFLMGNGKMLTKKDAQOLEYAAKQEPPLSLSDLDLRNN-TPLMSSRPHNPMVLP	159
Db	93	GGDAL-----PADSAGETADITYPLSLMTDLDDKNDRLGLGVREHNPMPLMP	139
Qy	200	IFMGKPKRSPNTPSPHSHEAR-OFTTFNEFRAPELKQVSVKAAEDLGTSDLMFGYTOQ	258
Db	140	LMYNNSPVYAGSPDTRGTGTVEKCGOQKRAETLKQVAFKSLAEDLFTKTAADLMFGYTQR	159
Qy	259	SHQOLF-N GKASRPPRRHNDIQPELFLTPQVYSDLPMWDKVMRTGMGAVHNSNGESAKLSR	317
Db	200	SDMOIYNGRKSAPFRNDYPELFLTPQVAKADLPFGGRLMLGAGFVHOSNGSGRESR	259
Qy	318	SMNRATYLMAGEMKMLMPMPRIKREKSGSGQPDNDPILOYGYGDVRFPLYOLENKS	377
Db	260	SMNRTYLMAGHEKCKLVIYIPVWYRATDQ-SGDK-NQNPDIADMGIDVAKLOTRLMDRQ	317
Qy	378	NISGTVRYNPRSGKALQLDVYVPLGKGISGTFQIFQSGSLIDYMHKATSEFGVGLMLN	437
Db	318	NVSYLRNPKXTGATGEAAVTTFPIKGLKGVVGRFGHGYGSLIDYMHKONGIGILMFN	377
Qy	438	DWMLG 442	
Db	378	DLDGI 382	
RESULT	2		
Q9J721			
ID	Q9J721	PRELIMINARY:	PRT: 409 AA.
AC	Q9J721		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	PUTATIVE PHOPHOLIPASE.		
GN	NMA2021.		
OS	Neisseria meningitidis (serogroup A).		
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.		
OX	NCBI_TaxID=65699;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;		
EX	MEDLINE-20222556; Pubmed-10761919;		
RA	Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,		
RA	Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,		
RA	David R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holroyd S.,		
RA	Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,		
RA	Randread M.A., Rutherford K.M., Simmonds M., Skelton J.,		
RA	Whitehead S., Spratt B.G., Bartell B.G.;		
RT	"Complete DNA sequence of a serogroup A strain of Neisseria		
RT	meningitidis 22491."		
RL	Nature 404:502-506(2000).		
DR	EMBL: AL162757; CAB85240.1; --		
DR	InterPro: IPR003187; Pfam:		
DR	Pfam: PF02253; Pfam: 1.		
Qy	Complete proteome.		
Qy	SEQUENCE 409 AA: 45862 MW: CD6585B064D01A41 CRC64;		

[illegible][illegible]

Query Match	15.8%	Score 373.5	DB 2	Length 306
Best Local Similarity	37.6%	Fred. No. 1.6e-24		
Matches 86	Conservative 43	Mismatches 91	Indels 9	Gaps 5
Qy	214	SHEAQTPEPNE-FRABELKFOYSVVKAAEDLWGTDSDLMFGYTOOSHQIFENGKSRPF	272	
Db	81	TSNHFLEKRETKQKDELFEKISLPLMRGLGNNSVLAASYTQASWFOLSNVDDSSPF	140	
Qy	273	RVHDPQPELFLQAPVUSDPMDGKRYMLMGAVHHNSNG--ESAKLSRSNNRAYLAGMEM	330	
Db	141	RETNERPQLAMKQOYSLPFGWITLQDVETGLNHOSNGDDAEKLSRNNRLYVRSAIK	200	
Qy	331	KNLVMPRIWGHIFEGSGSDPDNDPDLIDYGYGQVRF-LYOLENKSINISGTVRYNPRS	389	
Db	201	QMWTEVEIKPWWRIPEK---AKNDNDPDITKRYGHFEDLVAGLYUHHQFKLSG--HYNPIS	255	
Qy	390	GKGALQDLVYVPLGKISGYFOIFPGYGOSLDLYNHEATSFVGYGLMLND	438	
Db	256	NKGLEASYSYPTKNNIREYTYUNGYSGLDYQORIQRIGISGLNN	304	
RESULT	4			
Q9PMU8				
ID	Q9PMU8	PRELIMINARY:	PRT:	329 AA.
AC	Q9PMU8			
DT	01-OCT-2000	(TREMBLrel. 15	Created)	
DT	01-OCT-2000	(TREMBLrel. 15	Last sequence update)	
DT	01-JUN-2001	(TREMBLrel. 17	Last annotation update)	

```

DE PHOSPHOLIPASE A (EC 3.1.1.32).
PLDA OR C1351.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_taxid=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTC 11168;
MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtoyd S.,
RA Jorgensen K., Kariyasev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajadream M.A., Rutherford K.M., Van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
DR EMBL; AL139078; CAB73778.1; -
DR InterPro; IPR003187; PLA1.
DR Pfam; PF02253; PLA1; 1.
DR PRINTS; PR01486; PHPLIPASEA1.
KW Complete proteome.
SQ SEQUENCE 329 AA; 38880 MW; A32731F2B751AC44 CRC64;

Query Match 15.5%; Score 366; DB 2; Length 329;
Best Local Similarity 33.2%; Pred. No. 8.2e-24;
Matches 105; Conservative 48; Mismatches 125; Indels 38; Gaps 13;

QY 130 IKGRQVYVQETDPIFLMGNEKGLTKDAKOLEYAKQPTPLSPFDLRNNT--PLW 187
DB 43 LKNSVLSIQEONNSQATOTQNSTITKEKQ-----DSRLALANYLQENSPFLG 96
QY 188 SSRPHNRYVLPFMHCKPNRSPPTPSHARQFTPNFRABELFQYSVKKAEDLMGT 247
DB 97 IS-SYKKNYFLP-PAYSFNSLGVNNKSEA-----KFOLSVKRLRLENLGL 141
QY 248 DSDLMFGYTQSHQWTFNGKNSRPRRYADYQPELFTQPVY-SDLPMGDKRAMGCAVH 306
DB 142 DEKYIYAVTQTSWQIY--EHSSPFRFTNYQPEFFIDLPYLKDYEFFNNLR--VGIIH 196
QY 307 HSNESAK--LSRSMNAYILMAGMEKNLTYVPRMGRIFKEGSGSDPDNDLIDYGY 364
DB 197 ESNCKGDENLQSRSMNKIYVSTAILNKLEFVPRLMYRI---PENKDDDPALHLHWGN 253
QY 365 GDVRFYQLEKNSISGTVRYNPR--SGKALQLDYVYPL-GKGISGYFOIFOGYGSGLI 421
DB 254 FDVNLAV-LGDDYFIINLMLRNKLKFNHNKGAIOYDLDGINNGIYWLQYFNGYGESLI 312
QY 422 DYNHEATSPGVGLMLN 437
DB 313 DYNKHLQRLSTGFLIS 328

RESULT 5
Q9S1D7 PRELIMINARY; PRT; 292 AA.
ID Q9S1D7;
AC Q9S1D7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PHOSPHOLIPASE A.
GN PLDA.
OS Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=633;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YPIII PIB1;
RA Kariyasev A.V., Minzeler E.A., Williams K.J., Oyston P.C.,
RA Tiltball R.W., Wren B.W.;

```

```

RT "Biochip-based Signature-Tagged Mutagenesis: Identification and
RT characterisation of Y. pseudotuberculosis gene plda essential for
RT virulence in mice."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ245393; CAB51586.1; -
DR InterPro; IPR003187; PLA1.
DR Pfam; PF02253; PLA1; 1.
DR PRINTS; PR01486; PHPLIPASEA1.
KW Complete proteome.
SQ SEQUENCE 292 AA; 33758 MW; 8E712D908ACB6BA5 CRC64;

Query Match 15.3%; Score 360.5; DB 2; Length 292;
Best Local Similarity 39.2%; Pred. No. 2.1e-23;
Matches 83; Conservative 34; Mismatches 86; Indels 9; Gaps 5;

QY 229 ELKRFQVSVKKAEDLMGTDSDLMFGYTQSHQWTFNGKNSRPRRYADYQPELFTQPV- 287
DB 86 EVKFOLSLAFPIWIRGIDGNSLGLASYSQSRSMQVANSSESPRETYEQLFLAWSTD 145
QY 288 YSDLPMGDKRAMIGCAVHNSNGESAKLSRSMNRYILMAGMEKNLTYVPRMGRIFKEG 347
DB 146 YELAGW--TFREVERGFENHQSNGKADPTSRSMNRYITRYVMAQRGLETDLKPMYRIIPSD 203
QY 348 SGSDPDNDPLIDYGYGDVRFYQL-ENKNSISGTVRYNPRSGKALQLDYVYPLGKI 406
DB 204 S---KDDNPDLITKYNMGYRLKAVGALGDSVFSLDG--RYNNMTGIGAKMGMSYPIFKHV 258
QY 407 SGYFOIFQGYQSLIDYVNEATSPGVGLMLND 438
DB 259 RFTYQVFSYGESMIDYVNFROTGVGIMLND 290

RESULT 6
Q9L6N9 PRELIMINARY; PRT; 289 AA.
ID Q9L6N9;
AC Q9L6N9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PLDA PROTEIN.
GN PLDA.
OS Salmonella typhimurium LT2.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=99287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SGSC1412;
RA Washu.
RT "The Salmonella typhimurium Genome Sequencing Project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SGSC1412;
RA Waterston R.;
RA Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF333324; AAF33435.1; -
DR InterPro; IPR003187; PLA1.
DR Pfam; PF02253; PLA1; 1.
SQ SEQUENCE 289 AA; 32967 MW; DA97F5E1651C49C6 CRC64;

Query Match 15.3%; Score 360; DB 2; Length 289;
Best Local Similarity 39.7%; Pred. No. 2.3e-23;
Matches 85; Conservative 29; Mismatches 90; Indels 10; Gaps 4;

QY 226 RAPELKRFQVSVKKAEDLMGTDSDLMFGYTQSHQWTFNGKNSRPRRYADYQPELFTQ 285
DB 83 RKDEVKFOLSLAFPIWIRGIIIGLSPNSVLGASYSQSRSMQVANSSESPRETYEQLFLG 142
QY 286 PV-YSDLPMGDKRAMIGCAVHNSNGESAKLSRSMNRYILMAGMEKNLTYVPRMGRIF 344
DB 143 ATDVRFAQW--TLRDVEMGYHNSNGRSDPTSRSMNRLYTRLMANGMWLVEYKPMYVI- 199

```

```

QY 345 KGGSGQPDNDPILDYGYGVRLYOLENKSNSIGTVRYNPRSGALQLDVYPLGK 404
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 200 -----GSTDNDPITFYMGYQIKIYHL-GEAVLSAKQYMMNTGYGAEGVLSYPTK 253
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 405 GISGYFOIFQGYGOSLIDYNHEATSGVGLMND 438
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 254 HVRLYQVYSGYGESLIDYNFQTRVGVGLMND 287
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
Q924N8 PRELIMINARY: PRT: 292 AA.
ID 0924N8:
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, last sequence update)
DE 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE OUTER MEMBRANE PHOSPHOLIPASE A PRECURSOR (EC 3.1.1.32).
GN PLDA.
OS Enterobacter agglomerans.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94131966; PubMed=8300539;
RA Brok R.G., Brinkman E., van Boxtel R., Bekkers A.C., Verheij H.M.,
RA Tommassen J.;
RT "Molecular characterization of enterobacterial plda genes encoding
RT outer membrane phospholipase A.";
RT J. Bacteriol. 176:861-870(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120145; PubMed=9921577;
RA Brok R.G., Boots A.P., Dekker N., Verheij H.M., Tommassen J.;
RT "Sequence comparison of outer membrane phospholipases A: implications
RT for structure and for the catalytic mechanism.";
RT Res. Microbiol. 149:703-710(1998).
DR EMBL: AF034414; AAD03498.1; -.
DR InterPro: IPR003187; PLA1.
DR Pfam: PF02253; PLA1.
DR PRINTS: PR01486; PHPLIPASEA1.
DR Signal: Hydrolyase.
KM SIGNAL: 1 20 POTENTIAL.
FT CHAIN: 21 292 OUTER MEMBRANE PHOSPHOLIPASE A.
SQ SEQUENCE 292 AA; 33719 MW; B7516D093B2BEEA CRC64;

Query Match 15.1%; Score 356.5; DB 2; Length 292;
Best Local Similarity 40.1%; Pred. No. 47e-23;
Matches 85; Conservative 31; Mismatches 87; Indels 9; Gaps 5;

QY 229 ELKFOYSVKKAEDLMGTDSLMEGYTQOSHMOIFNGKNSRPFVHDYQPEITLQPV- 287
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 86 EVKFOIUSLGFPIWRIAGTADNSLLGASYTRSMWQASNSDESSPFRRTNEPQIFLAWMD 145
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 288 YSDLPMDGVVRMIGMGAHVHNSGESAKLSRSMNRAVLAAGMEKKLTVNPRJMGRTFKG 347
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 146 YELAGW--TFREVEFGYNHQSNGKADPTSRSDVRYTRLMAQRNLEIDLKFWYRIPESD 203
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 348 SOSOPDNDPDIIDYGYGVRLYOL-ENKSNSIGTVRYNPRSGKALQLDVYVPLGKI 406
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 204 S---KDDNDPDKMYKLYLKYALGESVFSIDG--RTNMNTGTGGAGMGSYPTTKHV 258
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 407 SGYFOIFQGYGOSLIDYNHEATSGVGLMND 438
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 259 RFTYQVFSGYGESMIDYNFQTRVGVGLMND 290
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
Q92349 PRELIMINARY: PRT: 297 AA.
ID 092349:
AC 032349:
AC 032349:

```

```

DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, last sequence update)
DE 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE PHOSPHOLIPASE A.
GN PLDA.
OS Campylobacter coli.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=195;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-UA585;
RX MEDLINE=97230284; PubMed=9119448;
RA Grant K.A., Belandria I., Dekker N., Richardson P.T., Park S.F.;
RT "Molecular characterization of plda, the structural gene for a
RT phospholipase A from Campylobacter coli, and its contribution to cell-
RT associated hemolysis.";
RT Infect. Immun. 65:1172-1180(1997).
DR EMBL: Y11031; CA871915.1; -.
DR InterPro: IPR003187; PLA1.
DR Pfam: PF02253; PLA1.
DR PRINTS: PR01486; PHPLIPASEA1.
SQ SEQUENCE 297 AA; 34998 MW; 04B54A7BCA3764CE CRC64;

Query Match 14.5%; Score 342.5; DB 2; Length 297;
Best Local Similarity 32.4%; Pred. No. 8e-22;
Matches 97; Conservative 37; Mismatches 108; Indels 57; Gaps 11;

QY 178 DLDNRNTPLMSSRPNHPVPLPIFMHGKRNRPNT-----P 213
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 16 DLKENNASLSIKKHETQ-----NTQKTPSTKEDSRALANYLGENSEFPNIGIS 66
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 214 SHEARQFP-----NEFRAPELKFOYSVKAABEDLMGTDSLMEGYTQOSHMOIF 264
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 67 SYKMVYFLFPAVSFGSLGGENKRTKMKFQSLKRLFEDLLGIGERYVGYQTSMWQ-- 124
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 265 NKNKNSRPFVHDYQPEITLQPV-YSDLPMDGVVRMIGMGAHVHNSGESAK--LSRSMNR 321
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 125 NYKHSSPFRETNYOPEFFDIDLHEDYKFLNNLR--VGLIHESGKGDENLESRSNMR 181
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 322 AYLMGMEKKNTLVNPRJMGRTFKGSGSOPDNDPDIIDYGYGVRLYOLENKSNSIG 381
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 182 IYASSFYLRFLFVPRIMYRI---PENSEDDNDPEITMYKMGFDIN-IGSLGNDYFIYL 237
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 382 TVRYND--RSGKALQLDVYPL-GKGISGYFOIFQGYGOSLIDYNHEATSGVGLMND 437
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 238 MLRNNLDFHDKGAVQVDIGYDIFDNGIYWIQYFNGYGSLLDYNKRLRISTAFLLIS 296
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
Q925241 PRELIMINARY: PRT: 355 AA.
ID 0925241:
AC 025241:
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, last sequence update)
DE 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE PHOSPHOLIPASE A1 PRECURSOR (DR-PHOSPHOLIPASE A).
GN HP0499.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalam H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,

```



```

QY 70 EHEPELYTALAKMTLMNCALNODIMPLACYDILVHGETPAVITKRSIRLDETIW- 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10 EHAKTLDITLTKRKLEIEDITPNOYTMYSIDRLVWG---GIMVDEIDITFDGIGRQ 66
QY 129 -----TIKGPQVYVQETDPIFLMGNEKGMILTKRKAQOLEVAAKQFTPL 173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 PGVNFLELERELGLINIGSPAKIVIDGTS---YEVGENEALVYGKAKALAS----- 116
QY 174 SLSPDLDRNNTPMLSSRPHPNPMVYLPIEMHGKPNRSPNTPSHEARQFTNEFRAPELKQ 233
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117 --SLDSAKPAKLYNAPAAVAEPFRITIQDDAIKAPLDGVTCKNRTICKYLVPEV--- 171
QY 234 VSVKKAADLNGTDSDLFGYT---QDSHMOIFNGKNRPRFRVNDYQEFITLQPVYSD 290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 ---VECTQ-----LSMGLTRLAEGSNW-----NSMPHTHRRMEVYF---YFD 209
QY 291 LPMDGKVRMIGGAVHNS-----NGESAKLSKSNRAYSIMAGEMENKLVMPRIMGRPK 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210 MAEDTIFHM-KGEPRHETNLMVHNEQAVISPSWS---IHTGVTKNTVA---IMGMT-- 260
QY 346 EESGSOQPDNDPILDYGYGDVR 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 -GENLTFDD---MDHIMLDIR 278

```

RESULT 12

```

Q9F216 PRELIMINARY; PRT; 824 AA.
AC 09F216;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CELULOSE.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxId=1409;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KSM-S237;
RX MEDLINE=21036886; PubMed=1193393;
RA Hakemada Y., Hatada Y., Koike K., Yoshimatsu T., Kawai S.,
RA Kobayashi T., Ito S.;
RT "Deduced amino acid sequence and possible catalytic residues of a
RT thermostable, alkaline cellulase from an alkaliphilic Bacillus
RT strain.";
RL Biosci. Biotechnol. Biochem. 64:2281-2289(2000).
DR EMBL: AB018420; BAB19360.1; -.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR Pfam: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
SQ SEQUENCE 824 AA; 91564 MW; 65FA940FE1D729B9 CRC64;

```

Query Match 4.6%; Score 109.5; DB 2; Length 824;

Best Local Similarity 19.7%; Pred. No. 0.76; Matches 91; Conservative 57; Mismatches 146; Indels 167; Gaps 27;

```

QY 30 PVAFADEVSKND-----LGDNELLIGVQATQASSTPTANPL-DEHEPELY-----TT 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25 PALAAEGNTRDNKHLGNDNVKRPSEAGALQLEVDGQMTLVDOHGEKIQLGKMSIH 84
QY 79 ALENKTMILNCAL-----NODIMRLACY-DLVHGETPAVITKRSIRLDETI-----TS 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 85 GLQWPELINDAVAKALANDMESNMIRLAMY---VGENGYATNPBELLKQRYIDGIELAI 140
QY 121 RLDERTI---WQT-----TIKGPQVYVQETDPI-----IF 146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 ENDWYVIVDMVHAPGPRDPVYAGAKDFREIALALYPNPHIYELANPESSNNNGAG 200
QY 147 IMGNEKGMILTKRKAQOLEVAAKQFTPLSLSPDLDRN---NTPLMSSRP-----H 192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 201 IPNNEBGKAVK-----EYADPIVEMLRKSGNADNIIIVGSPNMSQRPDLAANDPIDH 255
QY 193 NPMYVLPPIFMHGKPNRSPNTPSHEARQFTNEFRAPELK-----POVSVKVAEDLMGT 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 HTMYTVHYFTGSHAATESYPSSE-----TPNSERGVMSNTRVLAENGVAVFATE--WQT 308
QY 248 -----DSDLMFGYTQO-----SHMOIFNGKNS-----RPRV- 274
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 SQASGDGPGYFDEADVWIEFLNENNISMANWSLTN-KNEVSGAFPTPELGKSNATNLDPG 367
QY 275 --HDYQP-EIFLT-----QPV-----YSDLPV---DGKVRMIGGAVHNSGE 311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 368 PDHVAPELSTSGEYVRAKRGVNTPEPIDRTKYTVLMDFPDGTQKQ--GFGVNSDSPAK 425
QY 312 SAKLSRSNRAVLAGME-----WKNLTVMRPIWGR 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 426 ELIADVNEENTLTKSGLDVSDVSDQNFANARLSANGMK 466

```

RESULT 13

```

Q99241 PRELIMINARY; PRT; 821 AA.
AC Q99241;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4) (CELLULOSE) (ENDOGALACTANASE)
DE (CARBOXYMETHYL CELLULOSE).
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxId=1409;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KSM-64;
RA Sumitomo N., Ozaki K., Ito S.;
RL Submitted (MAY-1993) to the EMBL/Genbank/DBJ databases.
CC -1 CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
DR EMBL: M84963; AAA73189.1; -.
DR HSSP: O85465; IAS3.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR Pfam: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 821 AA; 90910 MW; 73D438FEF0B40B5C CRC64;

```

Query Match 4.6%; Score 109; DB 2; Length 821;

Best Local Similarity 19.7%; Pred. No. 0.83; Matches 91; Conservative 51; Mismatches 150; Indels 170; Gaps 24;

```

QY 30 PVAFADEVSKND-----LGDNELLIGVQATQASSTPTANPL-DEHEPELY-----TT 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24 PALAAEGNTRDNKHLGNDNVKRPSEAGALQLEVDGQMTLVDOHGEKIQLGKMSIH 83
QY 79 ALENKTMILNCAL-----NODIMRLACY-DLVHGETPAVITKRSIRLDETI-----TS 126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 GLQWPELINDAVAKALANDMESNMIRLAMYVGENGYASNPELKSRYIKGIDLAIENDM 143
QY 127 ---WQT-----TIKGPQVYVQETDPI-----IFLGN 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144 YVIVDMVHAPGPRDPVYAGAEFFRDIALALYPNPHIYELANPESSNNNGAGIIPN 203
QY 151 EKGMLTKRKAQOLEVAAKQFTPLSLSPDLDRN---NTPLMSSRP-----HNPY 196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 204 EBGWNAVK-----EYADPIVEMLRSGNADNIIIVGSPNMSQRPDLAANDPIDHHHY 258
QY 197 VLPPIF--MHGKPNRS--PNTPSHEARQFTNEFRAPELKFOVSVKVAEDLMGT----- 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 259 TVHFTYGTSHAATESYSPETPNSERGVMSNTRVA---LENGVAVFATE--WGTSGANG 312
QY 248 -----DSDLMFGYTQO-----SHMOIFNGKNS-----RPRVH----- 275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

[illegible]

RESULT 14

AD	063485	PRELIMINARY;	PRT;	602 AA.
AC	063485.			
DT	01-NOV-1996 (TReMBLrel. 01, Created)			
DT	01-NOV-1996 (TReMBLrel. 01, Last sequence update)			
DT	01-JUN-2001 (TReMBLrel. 17, Last annotation update)			
DE	C-RAP ACTIVATED ONCOGENE FUSION PROTEIN.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euteleota; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87172791; PubMed=5550433;			
RA	Isihkawa F., Takaku F., Nagao M., Sugimura T.;			
RT	"Rat c-rat oncogene activation by a rearrangement that produces a			
RT	fused protein.";			
RL	Mol. Cell. Biol. 7:1226-1232(1987).			
CC	-1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.			
DR	EMBL: M15428; AAA42002.1; "			
DR	InterPro: IPR000719; Euk_Pkinase.			
DR	InterPro: IPR002290; Ser_thr_kin_acctsite.			
DR	Pfam: PF00065; pkinase; 1.			
DR	SMART: SM00221; STYKC; 1.			
DR	PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.			
DR	PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.			
DR	PROSITE: PS00108; PROTEIN_KINASE_SF; 1.			
KW	ATP-binding; kinase; Oncogene; Serine/threonine-protein kinase;			
KW	transferase.			
QO	SEQUENCE	602 AA;	69209 MW;	88B/BFA90FFB02AC CRC64;

Query Match	4.58;	Score 106.5;	DB 11;	Length 602;
Best Local Similarity	19.08;	Pred. No. 0.88;		
Matches	92;	Conservative 74;	Mismatches 172;	Indels 145;
			Gaps	24;

0Y	23	QAQAVPVAAYVDEYKRSNDIGONNELLIGQSQTQSADTDPANPDENEHELPTTALEN	82
Db	81	ELEKINNNOVKVLETKNKELETAORNL--GIQSOFTRAKE---LEAEKRDILT--N	131
0Y	83	KTMLINCALMODIMRLKLCYDPLVHNGEPAVIAKRSIRIDETIMO--TIKGPQVVOE	140
Db	132	ERLSQVEYLTEEDYKRL--NEKLESNT--TKGELDKIDELQASDVYVYKREKLEOE	166
0Y	141	-----TTDPIFLMGNGK-----MLTKDAQOLEAAQOFPLS	174
Db	167	KELLHNONSMILETKYTKTDELALGRKKEGELLECLKTEKNKEEDAIRSHSESAPSA	246
0Y	175	LSFDDLRNNTPLWMSRRPNHMYULPIFNHGKPNRSPNTPSHAEAQOTPNNEFAELFOY	234
Db	247	LS--SSPNMLSTPGMSQRPTR-----VPAQREKAPGSGTOEKKIKIPRGRODSUYWEI	298
0Y	235	SVKYVAARDLWCTGSDLMFEGTQOSH-----OLFNGKNSRRPRVYDXOPEI-----	281
Db	239	EASEVMJLSTRIGSGS---FGTYVYIGKMGHDGVAIVILKYVDPTRPQDLAFREVAVALFKTR	355
0Y	262	-----FLVQ-----PVYSDL--PMDGKVRMI-----GMGAVNH	307
Db	356	HVNILLFNGYMTKDLIALVTOGCESSSLYKIKLHNOETRFQOFOLIDARQTAQSMYDILHA	415

```
OY      308 SNGEASAKTSRSMNRYLWAGEMK-----NLVMPRIINGRIKESSGQ-----351
          |   | : | : | : | : | : | : | : | : | : | : | : | : |
Db      416 KNIHRODKS--NNIFLEHGELTVIKIGDEGLATVKSrw-----SGSQOYEOPGTGSYLMM 466
          |   | : | : | : | : | : | : | : | : | : | : | : | : |
OY      352 -----PDDNPDLI--DYVGCGDVRFLFYOLENKSNISGTYRVNPRSGKALDLDPYPL 402
          |   | : | : | : | : | : | : | : | : | : | : | : | : |
Db      467 APEVTARMODNNPFOSQIVSYGIY--LYEL-----MIGELPIYSHNRDI-----TFMV 515
          |   | : | : | : | : | : | : | : | : | : | : | : | : |
OY      403 GKG 405
          |   | : |
Db      516 GRG 518
```

RESULT

ID	09UN32.	PRELIMINARY;	PR1;	798 AA.
AC	09UN32.			
DT	01-MAY-2000 (TREMBLrel. 13, Created).			
DT	01-MAY-2000 (TREMBLrel. 13, last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)			
DE	PPAR GAMMA COACTIVATOR-1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxId:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Larroy D., Vidal H., Andreelli F., Laville M., Langin D.;			
RT	"Cloning and mRNA tissue distribution of human PPARgamma coactivator 1."			
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF159714; AAD51615.1; -.			
DR	InterPro; IPR000504; RRM.			
DR	Pfam; PF00076; rrm.1.			
DR	PROSITE; PS50102; RRM; 1.			
DR	SMART; SM00360; RRM; 1.			
SO	SEQUENCE 798 AA; 91056 MW; F68F9768BD94E1P4 CRC64;			

Query Match 4.5%; Score 106; DB 4; Length 798;
 Best Local Similarity 24.1%; Pred. No. 1.5;
 Matches 93; Conservative 36; Mismatches 115; Indels 142; Gaps 27;

```

QY      3 VSLSTLTSLPCFALILQAQAVPNVAAYDEVKRSKNDGONELLIGVQSTQAST 62
      ::: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      89 LAVLETTEDSLDPV-----DEDGIPSPDALTD-----GDVTTDNE-----ASPSSMP 129
QY      63 D-TANPLDHEPELCTYAL--ENKTMILIN-CSALNODIMRLACYDVLVHGE-----T 110
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      130 DCTPPPOAEERPELSKLKLLAPANTQLSYNECSGLS-----TONHANHNHRTIN 179
QY      111 PAVIKTKRSIRLDETTIWOT-----IKGKPO-----VYQETTD-----PIFLMGNE 151
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      180 PAIVATEMS-----WENKAKSICQOQKQPRRPOSELLKYLTUNDDPHTKPRENRRSS 232
QY      152 KGMLT--KKDAKOLEYAKKQPTPLSLSTDLDRNNTPLMSSRNPHM-----YLP 199
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      233 RDKCTSKKSKSHTOOSOHQAKPTLLS-----PLTPESPNDPKGSPENKTIERTLS 285
QY      200 IEMHCKPMKRS--PNTPSHEAROTPEEFA--DELKFOV-----SVKKAEDLMGT--- 247
      : : : : : | | | | | | | | | | | | | | | | | | | | | |
Db      286 VELSTAGLTPTTTPPHKANO--DNPFASFKLKSSCKTVPYPPPSKKPRYSSESS-GIQGN 344
QY      248 -----DSDLMFGYTOOSHMOIFNG-----KNSRP-----FRVHDY-----QPEIFL 283
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      343 NSTKGKPPQSEL---YAQLSKSSVLTGGEHERKTKRPRLFGHDHYCQGSINSTEILIN 399
QY      284 -----TQPYVSDLPMDGKV 297
Db      400 ISQELQDSRQLENKDVSSD--WOGGI 423

```

Search completed: November 30, 2001, 14:26:27
Job time: 564 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:27:00 ; Search time 28.02 Seconds

(without alignments)
578.367 Million cell updates/sec

Title: US-09-787-083-4
Perfect score: 2360
Sequence: 1 MKVSLSTLTSLPCFAILA.....YNHEATSPGVGLMLNDMNGL 442

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	373	15.8	286	1 P37446	Klebsiella
2	358	15.2	289	1 P37442	salmonella
3	357	15.1	289	1 P00631	escherichia
4	321.5	13.6	289	1 P37447	proteus vul
5	109	4.6	800	1 GUN_BACSL	
6	94.5	4.0	397	1 GUN_PAEPO	
7	93	4.0	4273	1 PKSM_BACSU	
8	92.5	3.9	849	1 SRKE_BRAOL	
9	92	3.9	324	1 YL70_ARCFU	
10	92	3.9	696	1 SCKI_SCHPO	
11	92	3.9	888	1 YJHO_YEAST	
12	91.5	3.9	969	1 SACH_STRSL	
13	91	3.9	496	1 CATR_DICDI	
14	90.5	3.9	1474	1 AZMG_HUMAN	
15	90.5	3.8	467	1 INVO_MOUSE	
16	90	3.8	1158	1 RI14_HUMAN	
17	90	3.8	1788	1 YP72_CAEEL	
18	89.5	3.8	486	1 VGL4_BPMLS	
19	89.5	3.8	669	1 AVY1_ALTHA	
20	89.5	3.8	1478	1 BCK1_YEAST	
21	89.5	3.8	1840	1 SUIS_RAT	
22	89	3.8	1426	1 CURT_MOUSE	
23	88.5	3.8	825	1 GUN3_BACSA	
24	88.5	3.8	992	1 PH82_DICDI	
25	88	3.7	810	1 AMDM_YEAST	
26	88	3.7	1115	1 DP3A_BACSU	
27	88	3.7	1382	1 MET_RAT	
28	87.5	3.7	844	1 PAC_KLUCI	
29	87.5	3.7	1379	1 MET_MOUSE	
30	87	3.7	503	1 CD44_RAT	
31	87	3.7	666	1 PD14_RAT	
32	87	3.7	842	1 LPFC_SALTY	
33	87	3.7	882	1 HSSI_RAT	

34	86	3.6	882	1 HSSI_HUMAN	P52848 homo sapien
35	86	3.6	1087	1 XYNX_CLOTM	P38535 clostridium
36	85.5	3.6	353	1 DCUP_BACSU	P32395 bacillus su
37	85.5	3.6	790	1 SEIL_MOUSE	O92266 mus musculu
38	85.5	3.6	794	1 SEIL_HUMAN	O92266 mus musculu
39	85	3.6	403	1 P37_MYCHR	O92266 mus musculu
40	85	3.6	491	1 TY3H_PNASP	P15363 mycoplasma
41	85	3.6	560	1 DTXH_CORBE	P11982 phasianidae
42	85	3.6	828	1 BGAL_BRAOL	P00589 corynebacter
43	85	3.6	1184	1 ALAC_ARATH	P49676 brassica ol
44	85	3.6	2231	1 SEIL_YEAST	P57792 arabidopsis
45	84.5	3.6	478	1 DHGB_ACTICA	O00416 saccharomyc

ALIGNMENTS

RESULT 1	ID	STANDARD	PRT	286 AA.
PAL_KLEPN				
AC P37446:				
DT 01-OCT-1994 (Rel. 30, Created)				
DT 01-OCT-1994 (Rel. 30, Last sequence update)				
DT 01-NOV-1997 (Rel. 35, Last annotation update)				
DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DETERGENT-RESISTANT				
DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-				
DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).				
GN PLDA.				
OS Klebsiella pneumoniae.				
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC Klebsiella.				
OX NCBI_TaxID=573;				
RN [1]				
RP SEQUENCE FROM N.A.				
RX MEDLINE=94131966; PubMed=8300539;				
RA Brok R.G.P.M., Brinkman E., van Boxtel R., Bekkers A.C.A.P.,				
RA Verheij H.M., Tomassen J.;				
RT "Molecular characterization of enterobacterial plda genes encoding				
outer membrane phospholipase A.;"				
RL J. Bacteriol. 176:861-870(1994).				
CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE				
CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.				
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-				
CC PHOSPHOCHOLINE + A FATTY ACID ANION.				
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-				
CC PHOSPHOCHOLINE + A FATTY ACID ANION.				
CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.				
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES				
CC LOCATED THERE.				
CC This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC the European Bioinformatics Institute. There are no restrictions on its				
CC use by non-profit institutions as long as its content is in no way				
CC modified and this statement is not removed. Usage by and for commercial				
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC or send an email to license@isb-sib.ch).				
CC -----				
CC EMBL: X76901; CAA54223.1; -				
DR PIR: B36971; B36971.				
DR PIR: S40129; S40129.				
DR InterPro: IPR003187; PLA1.				
DR Pfam: PF02253; PLA1; 1.				
KW Hydrolase; Lipid degradation; Outer membrane; Signal; Calcium.				
FT CHAIN 1 20				
FT SIGNAL 1 20				
FT ACT SITE 21 286				
FT ACT SITE 161 161				
FT ACT SITE 161 161				
SO SEQUENCE 286 AA; 32544 MW; 3E39F863085108A3 CRC64;				

Query Match 15.88; Score 373; DB 1; Length 286;
Best local Similarity 42.18; Pred. No. 3.3e-23;
Matches 90; Conservative 27; Mismatches 87; Indels 10; Gaps 5;

```

OY      226 RAPBCEQSVSYVKAADMDJMGDSMLMFCYYTQQSHMOITNGKNSRFRHVDQPELEFLQ 285
Db      80 RDEVKFOLSTLSEPLMRGRGLGDNSSLGASTOKSMWOLNNSKESEAFRTYNPEOLFIEP 139
OY      286 PV-YSDLPDGCVRVMIGMGAHVHSNGESAKLSRSNMRAVLAMGMEWKNTVPRIWGRIE 344
Db      140 ANDYGFAGC--TLROIEMKYNDHNSGRDSPTRSNNRLYLARLAONGNMLVEYKPP---x 194
OY      345 KEGSGSQDDNPDLIDYGYGVREPLYOLENKSNISGTIVRYNPRSGKALQLDYVYPICK 404
Db      195 VVGs---RDDNPDPIKKMYGYRLKAYVOL-GEALLSAOGQYMNMTGGYGAEIYGSYPTRK 250
OY      405 GISGFOIFGCGGSLDIYNHNEATSPGVGLMLND 438
Db      251 HVRAVTQTISTGESLIDYNENQTRVGYGLMLND 284

RESULT 2
PAL_SALTLY
AC      P37442; STANDARD; PROT; 289 AA.
DT      01-OCT-1994 (Rel. 30, Created)
DT      01-OCT-1994 (Rel. 30, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DETERGENT-RESISTANT
DE      PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
DE      ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
GN      PLDA.
OS      Salmonella typhimurium.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Salmonella.
OX      NCBI_TaxID=602;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE 94131966; PubMed-8300539;
RA      Brok R.G.P.M., Brinkman E., van Boxtel R., Bekkers A.C.A.P.,
RA      Verheij H.M., Tommassen J.;
RT      "Molecular characterization of enterobacterial pldA genes encoding
RT      outer membrane phospholipase A."
RL      J. Bacteriol. 176:861-870(1994).
CC      -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
CC      A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
CC      -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-
CC      PHOSPHOCHOLINE + A FATTY ACID ANION.
CC      -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-
CC      PHOSPHOCHOLINE + A FATTY ACID ANION.
CC      -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
CC      -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
CC      LOCATED THERE.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
Cc      EMBL; X76900; CAAS4222.1; -.
DR      PIR; A36971; A36971.
DR      PIR; S40131; S40131.
DR      STGene; SG10302; plda.
DR      InterPro; IPR003187; PLA1.
DR      Pfam; PF02253; PLA1; 1.
KW      Hydrolase; Lipid degradation; Outer membrane; Signal; Calcium.
FT      SIGNAL 1 20
FT      CHAIN 21 289 BY SIMILARITY.
FT      ACT_SITE 164 164 PHOSPHOLIPASE A1.
FT      ACT_SITE 164 164 BY SIMILARITY.
SQ      SEQUENCE 289 AA; 33003 MW; D5ACFE5E33EF704DBA CRC64;

```

```

Best Local Similarity 39.3%; Pred. No. 5,6e-22;
Matches 84; Conservative 30; Mismatches 90; Indels 10; Gaps 4

OY 226 RAPELKFQVSVVVKAAEDLMGTDSLMEGYTQOOSHMOJFNGKNSRPFVHDXQPELFTQ 285
    | :|||:| : : :| :|||:| :|||:| :|||:| :|||:|
Db 83 RDEYKVFQSLAFLPMLRGILGPNVSLGASVYQKSMWQJNSKSESSPFRFTNEPOLFI6F 142
OY 286 PV-YDSLPMDCQVRVIMGCAVHNSGESAKLSRSNNRAYLMAEMKNTLVNPRIMGRTF 344
    | :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 143 ADYDFPAGV-"TLRHVEEMVGNHDSNGRSDPTSRTNRLYTRLMAENGMLVEKPPVYI- 199
OY 345 KEGSSQSPDNDNDITDYGXGVDVRFLOYENKSNISGTVNRPBSGKALOLDYVPLK 404
    ||||| | | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 200 -----GSTDDNPDITKRYMKYIQLKIGYHL-GERAVLSAKQJNMNTGYGAEVGLSTPYK 253
OY 405 GISGFOIFQGYGOSLIDYVNHNEATSPCYGLMND 438
    :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 254 HVLRYTQVYSGGESLIDYVNFQTRGVGVMLND 287

RESULT 3
PAL_ECOLI STANDARD; PRT; 289 AA.
AC P00631:
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DETERGENT-RESISTANT
DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
GN PLCA OR B3821 OR Z5342 OR ECS4751.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxId=562, 83334;
    [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85157492; PubMed=6397464;
RA Homma H., Kobayashi T., Chiba N., Karasawa K., Mizushima H., Kudo I.,
RA Inoue K., Ikeda H., Sekiguchi M., Nojima S.;
RT "The DNA sequence encoding plidA gene, the structural gene for
RT detergent-resistant phospholipase A of E. coli.";
RL J. Biochem. 96:1655-1664(1984).
    [2]
RN SEQUENCE FROM N.A.
RP STRAIN-K12 / MG1655;
RX MEDLINE=92358234; PubMed=1379743;
RA Daniels D.L., Plunkett G., Iii, Burland V.D., Blattner F.R.;
RT "Analysis of the Escherichia coli genome: DNA sequence of the region
RT from 84.5 to 86.5 minutes.";
RL Science 257:771-778(1992).
    [3]
RN STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G., Iii, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN SEQUENCE FROM N.A.
RP STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11205551;
RA Perna N.T., Plunkett G., Iii, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";

```

RL Nature 409:529-533(2001).
 [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-0157:H7 / RIMD 0509952;
 RA MEDLINE-21156231; PubMed-11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
 Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 Kuhara S., Shiba T., Hatori M., Shingawa H.;
 RT Complete genome sequence of enterohemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [6]
 RP SEQUENCE OF 30-289 FROM N.A.
 RC STRAIN-K12;
 RA MEDLINE-85003590; PubMed-6383820;
 RA de Geus P., Verheij H.M., Riegman N.H., Hoekstra W.P.M., de Haas G.H.;
 RT "The pro- and mature forms of the *E. coli* K-12 outer membrane
 RT phospholipase A are identical.";
 RL EMBO J. 3:1799-1802(1984).
 RN [7]
 RP SEQUENCE OF 174-289 FROM N.A.
 RC STRAIN-K12;
 RA MEDLINE-87115164; PubMed-3027506;
 RA Irino N., Nakayama K., Nakayama H.;
 RT "The recQ gene of *Escherichia coli* K12: primary structure and
 RT evidence for SOS regulation.";
 RL Mol. Gen. Genet. 205:298-304(1986).
 RN [8]
 RP MUTAGENESIS OF SER-172.
 RA MEDLINE-94131966; PubMed-8300539;
 RA Brok R.G.P.M., Brinkman E., van Bortel R., Bekkers A.C.A.P.,
 RA Verheij H.M., Tommassen J.;
 RT "Molecular characterization of enterobacterial plda genes encoding
 RT outer membrane phospholipase A.";
 RL J. Bacteriol. 176:861-870(1994).
 RN [9]
 RP ACTIVE SITE SER-164, AND PARTIAL SEQUENCE.
 RA MEDLINE-91249806; PubMed-2040286;
 RA Horrevorts A.J.G., Verheij H.M., de Haas G.H.;
 RT "Inactivation of *Escherichia coli* outer-membrane phospholipase A by
 RT the affinity label hexadecanesulfonyl fluoride. Evidence for an
 RT active-site serine.";
 RL Eur. J. Biochem. 198:247-253(1991).
 CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
 CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
 CC REQUIRED FOR EFFICIENT SECRETION OF BACTERIOCINS, SEEMS TO BE
 CC DORMANT IN NORMAL GROWING CELLS.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
 CC LOCATED THERE.
 CC -1- INDUCTION: BY MEMBRANE DAMAGING, FOR EXAMPLE, BY PHAGE-INDUCED
 CC LYSIS OR TEMPERATURE SHOCK.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X02143; AAA26081.1; -
 CC EMBL: M87049; AAA67617.1; -
 CC EMBL: AE000458; AAC76824.1; -
 CC EMBL: AE005613; AAG59017.1; -
 CC EMBL: AP002567; BAB38174.1; -
 CC EMBL: M30198; AAA24516.1; -
 CC PIR: A00771; PSECA.

DR PIR: A22133; PSECA.
 DR PIR: S30711; S30711.
 DR Ecogene; EG10738; PLDA.
 DR InterPro; IPR003187; PLA1.
 DR Pfam; PF02253; PLA1; 1.
 KW Hydrolase; Lipid degradation; Outer membrane; Signal; Calcium;
 KW Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 289 PHOSPHOLIPASE A1.
 FT ACT_SITE 164 164
 FT MUTAGEN 172 172 S->F: INACTIVE PROTEIN.
 FT CONFLICT 14 15 LP -> FA (IN REF. 2).
 FT CONFLICT 30 33 DAPA -> WTRQ (IN REF. 6).
 SQ SEQUENCE 289 AA; 33163 MW; A688AD32AA60F218 CRC64;
 Query Match 15.1%; Score 357; DB 1; Length 289;
 Best Local Similarity 39.3%; Pred. No. 6.7e-22;
 Matches 84; Conservative 29; Mismatches 91; Indels 10; Gaps 4;
 QY 226 RAPELRFQVSYKKAADLDNGTSDLFNGYTOOSHQDIFNGKNSRRPRVDYDFEFLTQ 285
 DB 83 RKDEVKFQLSLAPFLMWSGLIGPNSVLGASYTKQKSMQLNSSESPRENTYEPQLGIF 142
 QY 286 PV-YSDLPMQKVMIGMAVHHSNGESAKLSRSNMNAVYLMAGEMKRLVYPRIMGRIF 344
 DB 143 ATDYRFAGW--TLRDVEKGNHDSNGSDPTSRSMNLRYLRMAENGWMLVEYKPMIVV- 199
 QY 345 KEGSGQDDNPDIIDYGYGDVRFYQLENKSNISGTVRYNPRSGALQLDYVPLGK 404
 DB 200 -----GNTDDNPDTIKYMGYYQLKIGYHL-GDAVLSAKGQNMWMTGAGALGLSTPYTK 253
 QY 405 GISGFQIFQGYGQSLDYDNEHATSPGVGLMD 438
 DB 254 HVRLYTQVYSGYGSLLIDYFNQTRVGWGLND 287
 RESULT 4
 PAL_PROVU STANDARD; PRT; 289 AA.
 ID PAL_PROVU
 AC P37447;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, last sequence update)
 DT 01-NOV-1997 (Rel. 35, last annotation update)
 DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DEFERMENT-RESISTANT
 DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
 DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
 GN PLDA.
 OS Proteus vulgaris.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Proteus.
 OX NCBI_TaxID=585;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-94131966; PubMed-8300539;
 RA Brok R.G.P.M., Brinkman E., van Bortel R., Bekkers A.C.A.P.,
 RA Verheij H.M., Tommassen J.;
 RT "Molecular characterization of enterobacterial plda genes encoding
 RT outer membrane phospholipase A.";
 RL J. Bacteriol. 176:861-870(1994).
 CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
 CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
 CC LOCATED THERE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its


```
RT "Molecular cloning, expression, and characterization of
RT endo-beta-1,4-glucanase genes from Bacillus polymyxa and Bacillus
RT circulans".
RL J. Bacteriol. 172:1576-1586(1990).
CC -1- CATALYTIC ACTIVITY: ENDOPOLYMEROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M33791; AAA22631.1; -.
CC DR PIR: A35136; A35136.
CC DR HSSP: P54583; 1ECE.
CC DR InterPro: IPR001547; Glyco_hydro_F5.
CC DR Pfam: PF00150; cellulase_1.
CC DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC KW Cellulose degradation; Hydrolase; Glycosidase.
CC FT ACT_SITE 194 194 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 317 317 NUCLEOPHILE (BY SIMILARITY).
CC SO SEQUENCE 397 AA; 44357 MW; B9C2E802C04F0A2A CRC64;

Query Match 4.0%; Score 94.5; DB 1; Length 397;
Best Local Similarity 25.7%; Pred. No. 2.7;
Matches 46; Conservative 21; Mismatches 61; Indels 51; Gaps 10;

QY 301 GMAVHNSNESAKLSNMRAVLMGEMKNLTVMRIGR-----IFKRG----- 347
DB 43 GNRIVBSGKEAFNGLNW-----FLETPNPT-LHGLMSRSDMDLDQYKKEGYMLIR 95
QY 348 -----SGSOPD-----NPDILDYGVGVFLYQLEKNSINISGV-----RYNRP 389
DB 96 LPSYNGLFSSSRDSTIDYHKNDLV---GLNPQIMDKLEKAGQIGIQLIDRHRRPS 152
QY 390 GKALQLDIV--YPLGKISGYFQIFQGY-----GSLIDVNHETSEFGVGLMLNDW 439
DB 153 G-GSELMWTSQYPERSRMISDMKMLADRYKNPVTIGADLHNEPHGQASMGSTGNASTDW 210

RESULT 7
PKSM_BACSU STANDARD; PRT; 4273 AA.
AC P40872; O31781;
DT 01-FEB-1995 (Rel. 31, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE POLYKETIDE SYNTHASE PKSM.
DE PKSM OR PKSY.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBL_taxid=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-1763 FROM N.A.
RC STRAIN=168 / PB1424;
RA Tognoni A., Grandi G.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC SECONDARY METABOLISM.
CC -1- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES
```

```
CC (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z99113; CAB13603.1; -.
CC DR EMBL: Z35133; CAA84505.1; -.
CC DR Subtilist; BG10931; PKSM.
CC DR InterPro: IPR000794; ketoacyl_syn.
CC DR InterPro: IPR001601; Meth-transf.
CC DR InterPro: IPR003880; Phosphopant_attach.
CC DR InterPro: IPR000051; SAM_bind.
CC DR Pfam: PF00550; PP-binding; 1.
CC DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 2.
CC DR PROSITE: PS00075; ACP_DOMAIN; 4.
CC DR TRANSFERASE; Acyltransferase; Multifunctional enzyme; Repeat; Complete proteome.
CC KW Phosphopantetheine;
CC FT DOMAIN 295 364 ACRYL CARRIER (ACP) 1.
CC FT DOMAIN 396 834 BETA-KETOACYL SYNTHASE 1.
CC FT DOMAIN 2190 2258 ACRYL CARRIER (ACP) 2.
CC FT DOMAIN 2322 2737 BETA-KETOACYL SYNTHASE 2.
CC FT DOMAIN 3532 3947 BETA-KETOACYL SYNTHASE 3.
CC FT DOMAIN 3410 3483 ACRYL CARRIER (ACP) 3.
CC FT DOMAIN 4140 4209 ACRYL CARRIER (ACP) 4.
CC FT BINDING 327 327 PHOSPHOPANTETHEINE (POTENTIAL).
CC FT BINDING 2222 2222 PHOSPHOPANTETHEINE (POTENTIAL).
CC FT ACT_SITE 2476 2476 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
CC FT BINDING 3446 3446 PHOSPHOPANTETHEINE (POTENTIAL).
CC FT ACT_SITE 3690 3690 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
CC FT BINDING 4172 4172 E -> V (IN REF. 2).
CC FT CONFLICT 103 103 O -> E (IN REF. 2).
CC FT CONFLICT 276 276 T -> S (IN REF. 2).
CC FT CONFLICT 289 289
CC SO SEQUENCE 4273 AA; 477459 MW; 3BBFC1A250AEB5A CRC64;

Query Match 3.9%; Score 93; DB 1; Length 4273;
Best Local Similarity 19.2%; Pred. No. 97;
Matches 102; Conservative 79; Mismatches 163; Indels 188; Gaps 29;

QY 14 PCFAIILAQQAAY-PNPVAFVDE-----VRSKN-----DLGQ 45
DB 3597 PLFQIIPKREASMDPRQIFLEAMHTFEDAGYMGDRIRKSGGVYVGESEYALUTG 3656
QY 46 DNELLIGVQSKTASDTNPANPDEHEBELYTTALEKTKMLI---NCSALNODIMRLACY 102
DB 3657 DTDYINGTONTATLSAR--IAYALDLKGNVMTLTACSSGVAIHQAQSAALRQG---DCE 3710
QY 103 DTLVHGEPNAVYIKTKRSIRLDETIMQTIKGPVYVQETTPPIFGMNEKGLTKKDAQ 162
DB 3711 MALAAGYTLNI-----SHMSFELIRAEMLSPNGCKVYDDQANG 3750
QY 163 L-----EYARQFTPLSLSPDLRNNTPLMSSRPNPVYLPFEHGRPN--RSPNPSHE 216
DB 3751 LVPGEAAVAVLKPLSKAIE-DKDH-----YGCIKASGVNDGKTNGITARN-PFSQ 3801
QY 217 ARQ-----FTPREFAPBELKQV-----VKVAAEDLMGTSDLMFGYTGQSHWQ 262
DB 3802 AELLIENTYKNEINPLDIDQYMAHSTGSLGDPLEVALTSVFSK-----YTKQKQFC 3854
QY 263 IFNGKNSRPFRVHDYQPEIFLQPVYSDLPMDGKVRNIGM-----GAVHHSNGES- 312
DB 3855 MIS--SIKPLIGHFAA-----SGTVALISHLMAKNKIITATHHCSENP 3898
QY 313 -----AKLSRSNRAVLMAGEMKNLTVMRIGRIFKRG-SGSO----- 351
```

```

Db 3899 YIPKESPVLCENKNSWIK-----KN--QKPRM-GTISTGISTGNHAAVIEEYI 3946
QY 352 PDDNPDLITYGYGVY-----RELYOLENKSNIS-GTVRYNPRSGKAL 394
Db 3947 PDDQSTQHQSGSPQIFVISAQNDRLQDQACRMIVALEQNHLSPDVAAYTLQVGRKAM 4006
QY 395 QL-----DYYVPL-GRGISGYFQ-IFGYSQSLDYHNEA 427
Db 4007 EARLAIVANNQOLYRKLEKVEYKAMKNGVSGQQRSLYGTGILEODEA 4058

RESULT 8
SRK6_BRAOL STANDARD: PRT: 849 AA.
AC 009092:
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PUTATIVE SERINE/THREONINE KINASE RECEPTOR PRECURSOR (EC 2.7.1.37)
DE (S-RECEPTOR KINASE) (SRK).
CN SRK6.
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3712;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, S656; TISSUE=Stigma;
RX MEDLINE=92020942; PubMed=1681543;
RA Stein J.C., Howlett B., Boyes D.C., Nasrallah M.E.;
RT "Molecular cloning of a putative receptor protein kinase gene encoded
at the self-incompatibility locus of Brassica oleracea.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:8816-8820(1991).
CC -1- FUNCTION: INVOLVED IN SPOROPHYTIC SELF-INCOMPATIBILITY SYSTEM
(FERTILIZATION), PROBABLY ACTING IN COMBINATION WITH S-LOCUS-
SPECIFIC GLYCOPROTEINS. INTERACTION WITH A LIGAND IN THE
EXTRACELLULAR DOMAIN TRIGGERS THE PROTEIN KINASE ACTIVITY OF THE
CYTOPLASMIC DOMAIN.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN - ADP + A PHOSPHOPROTEIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY IN THE PISTIL AND ANTHR.
CC -1- POLYMORPHISM: THERE ARE A NUMBER OF DIFFERENT S ALLELES IN
B. OLERACEA, POSSIBLY PROVIDING THE RECOGNITION SPECIFICITY.
CC -1- SIMILARITY: THE EXTRACELLULAR DOMAIN IS SIMILAR TO S-LOCUS
GLYCOPROTEINS OF BRASSICA, WHILE THE INTRACELLULAR DOMAIN IS
A SER/THR-PROTEIN KINASE RELATED TO RAF KINASES.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M76647; AAA33000.1; ALT_TERM.
DR HSBP: P11362; Ileg.
DR InterPro: IPR001480; B_lectin.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR003609; Pan_apk.
DR InterPro: IPR002290; Ser_thr_kin_actsite.
DR InterPro: IPR000858; Slocus_glycop.
DR Pfam: PF00069; pkinase.1.
DR Pfam: PF00954; S_locus_glycop.1.
DR SMART: SM00108; B_lectin.1.
DR SMART: SM00473; PAN_AP.1.
DR SMART: SM00221; STYK.1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP.1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST.1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.
KW transferase; Serine/threonine-protein kinase; signal; ATP-binding;

```

```

KM Transmembrane; Receptor; Glycoprotein; Self-incompatibility.
FT SIGNAL 1 32
FT CHAIN 33 849
FT DOMAIN 33 446
FT TRAMEM 447 466
FT DOMAIN 467 849
FT DOMAIN 528 779
FT NP_BIND 534 542
FT BINDING 556 556
FT ACT_SITE 653 653
FT CARBOHYD 47 47
FT CARBOHYD 120 120
FT CARBOHYD 196 196
FT CARBOHYD 260 260
FT CARBOHYD 314 314
FT CARBOHYD 389 389
FT CARBOHYD 442 442
SQ SEQUENCE 849 AA; 97231 MW; 7E156059EDDF4370 CRC64;

Query Match 3.9%; Score 92.5; DB 1; Length 849;
Best Local Similarity 21.9%; Pred. No. 11;
Matches 78; Conservative 50; Mismatches 107; Indels 121; Gaps 24;

QY 105 LVHGETPAVYIKRSIRLDETWTQIKGPKQVYQETDPIFGM---NEK---GMLTK 157
Db 24 LIHPALSTIYINT-----LSTESTLITSNKTLV---SPGISIEYGFRTSRWYGLMWYK 75
QY 158 K-DARQLEVAKQFTPLISFP-----LDRNNTPLSSRPHNMVYLPFIMHGK 205
Db 76 KVSDDRTYVAVANRNDPLSNALGTILSGNNLVLDHSNKPVMWTN-----LTRGN 125
QY 206 PNRSP-----NTPSHARQPT-----PNEFRAPELKFOYSVYKAAEDLM 245
Db 126 -ERSPVVAVELLANGNVNMDSSNMDASEYLWQSPDYPTDTLLPEKKGIVNLT----- 177
QY 246 GTDSDLMEGYTQSHWQTFNGKNSRPFVHYD-----PEFLTQ---PYVSDLPMDCK 296
Db 178 GLNRFL-----TSWRSSDDPSGDNF---SYKLTQSLPERYLSRENFPMHRSQPMWG- 226
QY 297 VMIMGAVH-----HSGESAKLSRSMNRAY-----LMAGMEKNLTVP--RI 339
Db 227 IFFSGIPEQKLSYVAVNFENNEVATTFKMTNNSPRLTIEGFGRLTTPWSIRI 286
QY 340 WGRIFKESGSGQDDNPDLIDY---YGYGVRELYOLENKSNIQTVR-YNPRS 389
Db 287 WNRFW-----SSPVD-PQCDTYIMCGPYAYCDV-----NTSPVCNCIOGFNPN 329

RESULT 9
ID YL70_ARCFU STANDARD: PRT: 324 AA.
AC 028112:
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHEITICAL PROTEIN AF2170.
GN AF2170.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klein H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervilavege A.R., Graham D.E., Kyripides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodok A., Zhou L.,

```


OY 313 AKLSRSMNRAVIMAG---MEMKNLTVMPRING---RIKESGSGSPDDNDPILDIYY---G 363
 DB 458 ATTNFCGTTTELAEVLEEDKGYKQVDFMSLGVLFEMCCGMSFPYAPDVQMYRNA 517
 OY 364 YGDVRF---LYOLEKNSINSGTVRYNPRSGKAL 394
 DB 518 FGKVRFPKGVLSSEGRSFVRLNPNHRLGAV 551

RESULT 11

YJH0_YEAST
 ID YJH0_YEAST STANDARD: PRT: 888 AA.
 AC P40361.
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HYPOTHEETICAL 104.3 KDA PROTEIN IN SMC3-MRP18 INTERGENIC REGION.
 GN YJ070C OR J1095 OR HRD888.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C.
 RX MEDLINE-95282514; PubMed-7762302;
 RA Vandenbol M., Durand P., Dion C., Portetelle D., Hilger F.;
 RT "Sequence of a 17.1 kb DNA fragment from chromosome X of
 RT Saccharomyces cerevisiae includes the mitochondrial ribosomal protein
 RT L8.";
 RL Yeast 11:57-60(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA Sor F.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE ADENOSINE AND AMP DEAMINASES FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; 234288; CAAB4052.1; -
 DR EMBL; 249345; CAAB9362.1; -
 DR EMBL; 249344; CAAB9361.1; -
 DR EMBL; X88851; CAAB3309.1; -
 DR PIR; S47120; S47120.
 DR SGD; S0003606; YJL070C.
 DR InterPro; IPR001365; A_deaminase.
 DR Pfam; PF00962; A_deaminase.1.
 KW Hypothetical protein; Hydrolase.
 SQ SEQUENCE 888 AA; 104263 MW; DB31A8086224114D CRC64;

Query Match 3.9%; Score 92; DB 1; Length 888;
 Best Local Similarity 22.1%; Pred. No. 13;
 Matches 66; Conservative 36; Mismatches 126; Indels 70; Gaps 15;

OY 128 QTIKGPQVVOE---TTDPIFLMGNEKMLTKKDA-----KOLEYAAKOFPLSL 176
 DB 155 VAPSTSTASBA---DTH-TEVDLKVSNSAANNAISKLGRKISIVEENMTSDQIV-A 209
 OY 109 ETPAVIKTKRSRIDETIMQITIGK-PQVVOEETTPDPIFLMGNEKMLTKKDAKOLEYAA 167
 DB 210 LTEETIKALKNVDFSD---DAIKGTISLITRNLDIV-----ASFLKODSKLAVPYF 259
 OY 168 KQTPPLSL-SFD-----LDRNNTPLWSSRPHPM-----YVLPIMHGKPNRSP 210
 DB 260 KADTLINNPANNTDAQIMKKEEDVDMSWVQDAKSGVNSNMWGYOLVIMAGAPNNKS 319
 OY 122 FINPF---ELVSQMRKRITIAASKODGISININDTEKMFLLYKPLPKTKFREDDK-----R 173

OY 274 VHDYOPEIFLQPVYSDLPMDGKYRMIGCAVHHNSGCSAKLSRSMNRAVIMAGEMKNL 333
 DB 174 FQD-----PSDDLNDSDSTGCGAATPHRRGYYPPSYFTHHYYTTSGLKAGKN 224
 OY 334 TVMPRIWRIKESGSGSPDDNDPILDIYYGVDVFFLYOLEKNSINSGTVRYNPRSGK 391
 DB 225 IKVP-YTGEYF-----DLEDY---KKQYIYHLSNQENTONPL--SPYSSK 263

RESULT 12

SACB_STRSL
 ID SACB_STRSL STANDARD: PRT: 969 AA.
 AC Q05242.
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE LEVANSUCRASE PRECURSOR (EC 2.4.1.10) (BETA-D-FRUCTOFURANOSYL
 DE TRANSFERASE) (SUCROSE 6-FRUCTOSYL TRANSFERASE).
 GN FTF.
 OS Streptococcus salivarius.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1304;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 25975;
 RX MEDLINE-93322332; PubMed-8331080;
 RA Ratham C., Giffard P.M., Jacques N.A.;
 RT "The cell-bound fructosyltransferase of Streptococcus salivarius: the
 RT cardoxyl terminus specifies attachment in a Streptococcus gordonii
 RT model system."
 RL J. Bacteriol. 175:4520-4527(1993).
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) =
 CC GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
 CC ALSO ACT AS FRUCTOSYL ACCEPTORS).
 CC
 CC -1- SUBCELLULAR LOCATION: SECRETED. CELL-WALL BOUND.
 CC
 CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; I08445; AAA71925.1; -
 DR InterPro; IPR003469; Glyco_hydro_68.
 DR Pfam; PF02435; Glyco_hydro_68.1.
 KW Transferase; Glycosyltransferase; signal; Cell wall.
 FT SIGNAL 1 ?
 FT CHAIN ? 969 LEVANSUCRASE.
 SQ SEQUENCE 969 AA; 103983 MW; D389B5B32ACF735A CRC64;

Query Match 3.9%; Score 91.5; DB 1; Length 969;
 Best Local Similarity 19.9%; Pred. No. 16;
 Matches 85; Conservative 54; Mismatches 161; Indels 127; Gaps 20;

OY 53 VQATQASASTQANPLDHEBELYTTALEKNTMLINCALMODIMRLA---CYTTLVHG 108
 DB 155 VAPSTSTASBA---DTH-TEVDLKVSNSAANNAISKLGRKISIVEENMTSDQIV-A 209
 OY 109 ETPAVIKTKRSRIDETIMQITIGK-PQVVOEETTPDPIFLMGNEKMLTKKDAKOLEYAA 167
 DB 210 LTEETIKALKNVDFSD---DAIKGTISLITRNLDIV-----ASFLKODSKLAVPYF 259
 OY 168 KQTPPLSL-SFD-----LDRNNTPLWSSRPHPM-----YVLPIMHGKPNRSP 210
 DB 260 KADTLINNPANNTDAQIMKKEEDVDMSWVQDAKSGVNSNMWGYOLVIMAGAPNNKS 319
 OY 211 NTPSHARQTFPNEFRAPDELKFOYSVKYKAAEDLMDTSDLIMFGYTOOSHWOIFNGKNSR 270

```
Db 320 NH1LLYRKYGDNDF-----THW-----KNAG 341
QY 271 PFRVHDYQPIFLTPYSDLPMDGKYRMIGKAV--HHSNGESAKISRSMMNR---AYLM 325
Db 342 P-----IFGYMNMLEDQOQMSGATVNSDSIOLYTKNDTSGGKLMWOLASATLN 392
QY 326 AGMEMKLTVMPLMGIRFKEGSG-----SQP-----DDNPDILDYGYGD--- 366
Db 393 LAVENDEYVAKSVENDHILFEGGDNHYHQSTPKMSTFDDDNHNDGNDPRDNTCLRDPHI 452
QY 367 -----VRFI-----VOLENK-----SNISGVYRNPGRSGALQDLYVPLGKIS 407
Db 453 IEDNGSRYLIFESNTGDENYQGEKQIYKMSYNGGDAFNILKSPFLYNNHVLNLSAWAN 512
QY 408 GYFOFO 414
Db 513 GSIGILK 519
```

RESULT 13

```
CATAL_DICDI STANDARD; PRT; 496 AA.
ID CATA_DICDI
AC 077229;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE CATALASE (EC 1.11.1.6).
GN CATA OR CAT.
OS Dictyostellium discoideum (Slime mold).
OC Eukaryota; Metazoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RA Foote C., Alexander H., Alexander S.;
RA Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: OCCURS IN ALMOST ALL AEROBICALLY RESPIRING ORGANISMS AND
SERVES TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN
PEROXIDE.
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) - O(2) + 2 H(2)O.
CC -1- COFACTOR: HEME GROUP (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: PEROXISOMAL (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CATALASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF090443; AAC36743.1; -.
DR HSSP: P00432; 7CAT.
DR DICTYDB: DD07272; CATA.
DR InterPro: IPR002226; Catalase.
DR Pfam: PF00199; Catalase; 1.
DR PRINTS: PR00067; CATALASE.
DR PRODOM: PD000510; Catalase; 1.
DR PROSITE: PS00437; CATALASE_1; 1.
DR PROSITE: PS00438; CATALASE_2; 1.
KM Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
KM Peroxisome.
FT ACT_SITE 54
FT ACT_SITE 128
FT BINDING 338
FT SITE 494
SO SEQUENCE 496 AA; 55683 MW; 683318B8FBAFDE6 CRC64;
```

Query Match 3.9%; Score 91; DB 1; Length 496;
Best local Similarity 23.5%; Pred. No. 7;

```
Matches 56; Conservative 32; Mismatches 94; Indels 56; Gaps 12;
QY 61 STQANPLDEHEPELTYTALENKTNLNGSALNODINRIACDYDTLVGEPAYIKTRSI 120
Db 7 TTSSGSPIDNN-----LNSMTAGVNGPILIDFTLI--DLAHPDRE----- 46
QY 121 RLDEETIMQITIKRPQVYQETTD-----PIF-----LMGNEKML-T 156
Db 47 RIERPVYHA-KGAGAHGYFEVTSDDYKCKAKFLNKYGRKTFLETFRFSYVSGKSSDS 105
QY 157 KDAKOLEYAAKQFTPLSLFDLDNRNTPLMSSRPHPMYVLPFMHGKPNRSPPSHE 216
Db 106 ERDPR--GFAVKFYTEEG-NEDWVGNNTPVPEFIRDPsk---FPDFHTQ-KRPQFNCKD 158
QY 217 AKQFTPNFAPRLKFOYSVKVAEDLMGDSLMGCTQOSHWQIFNGNSPPFY 274
Db 159 PNMFDFLGQTPSTHGVSLFSDR---GTPKSYRHHGFSHTLKFVNAQGRPYW 212
```

RESULT 14

```
A2MG_HUMAN STANDARD; PRT; 1474 AA.
ID A2MG_HUMAN
AC P01023;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE ALPHA-2-MACROGLOBULIN PRECURSOR (ALPHA-2-M).
GN A2M.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85190481; Pubmed=2581245;
RA Kan C.-C., Solomon E., Belt K.T., Chain A.C., Hiorns L.R., Fey G.H.;
RT "Nucleotide sequence of cDNA encoding human alpha 2-macroglobulin and
RT assignment of the chromosomal locus".
RT Proc. Natl. Acad. Sci. U.S.A. 82:2282-2286(1985).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT HIS-704.
RX MEDLINE=92246939; Pubmed=1374237;
RA Mathijs G., Devriendt K., Cassiman J.-J., van den Berghe H.,
RA Marynen P.;
RT "Structure of the human alpha-2-macroglobulin gene and its promotor.".
RT Biochem. Biophys. Res. Commun. 184:596-603(1992).
RN [3]
RP SEQUENCE OF 24-1474.
RX MEDLINE=84239807; Pubmed=6203908;
RA Sottrop-Jensen L., Stepanik T.M., Kristensen T., Wierzbicki D.M.,
RA Jones C.M., Loenblad P.B., Magnusson S., Petersen T.E.;
RT "Primary structure of human alpha 2-macroglobulin. V. The complete
RT structure.".
RT J. Biol. Chem. 259:8318-8327(1984).
RN [4]
RP ERRATUM.
RA Sottrop-Jensen L., Stepanik T.M., Kristensen T., Wierzbicki D.M.,
RA Jones C.M., Loenblad P.B., Magnusson S., Petersen T.E.;
RA J. Biol. Chem. 260:6500-6500(1985).
RN [5]
RP INHIBITORY SITE.
RX MEDLINE=84030513; Pubmed=6195065;
RA Virca G.D., Salvesen G.S., Travis J.;
RT "Human neutrophil elastase and cathepsin G cleavage sites in the bait
RT region of alpha 2-macroglobulin. Proposed structural limits of the
RT Hoppe-Seyler's Z. Physiol. Chem. 364:1297-1302(1983).
RN [6]
RP INHIBITORY SITE.
RX MEDLINE=81212827; Pubmed=6165619;
RA Sottrop-Jensen L., Loenblad P.B., Stepanik T.M., Petersen T.E.,
RA Magnusson S., Joernvall H.;
RT "Primary structure of the 'bait' region for proteinases in alpha 2-
```



```
Db 217 H----PFTVEEFLVLPKFEVQVTPKIIITILEEENMVSGLTGYKPYVGHVTSICRKY 273
OY 240 AEDLMCTDSDLMFGYIQOSHMOJFNCK-NSRPRVHDYOPEITLTPVYSDLPMDGKVR 298
Db 274 DASCHGEDS-----QAFCEKFSQOLNSHGCFYQOVKKYFOLRKREKEMKLTAEQ 325
OY 299 MIGCAVHNSNG-ESAKTSSMNR-----AYLMAGMEKNLTVMPRIMORI-FKGGSG 349
Db 326 IOEGGYVELTGRKSSSEITITIKLSFVKYDSHRCGIPF-----FGQVRLVDGKG 376
OY 350 SOPDNDPIL-----DYGC-----YGDVRFYOLENKSNSISCT-----VRYNPRS 389
Db 377 -VPIPKVIFIRGNEANYSNATDEHGLVQFSI---NTTNVMTSLVRYNYNDRS 429

RESULT 15
INVO_MOUSE STANDARD; PRT; 467 AA.
ID INVO_MOUSE
AC P48997;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE INVOLUCRIN.
GN IYL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIH SWISS;
RX MEDLINE=94104476; PubMed=8277848;
RA Djian P., Phillips M., Easley K., Huang E., Simon M., Rice R.H.,
RA Green H.;
RT "The involucrin genes of the mouse and the rat: study of their shared
RT repeats.";
RL Mol. Biol. Evol. 10:1136-1149(1993).
CC -1- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS
CC IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO
CC MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE
CC FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC -1- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND
CC OTHER STRATIFIED SQUAMOUS EPITHELIA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; L28819; AAA39330.1; -.
CC DR MGD; MG1:96626; IYL.
CC DR InterPro; IPR002360; Involutrin.
CC DR PROSITE; PS00795; INVOLUCRIN; 1.
CC KM Keratinocyte: Repeat
CC SO SEQUENCE 467 AA; 54919 MW; 603E1E51B435737D CRC64;

Query Match 3.8%; Score 90.5; DB 1; Length 467;
Best Local Similarity 19.8%; Pred. No. 7.1;
Matches 62; Conservative 44; Mismatches 128; Indels 79; Gaps 14;

OY 134 PUVVY-----OETDPIFLMG-----NEKGMLTKKDAKOLEYAKQFTPLSLFDDRN 182
Db 175 POELHLRQHOEKILODPRLNLGQOQKTPREQKILPGEQELHLGQRHQRQEPQDELHLGOK 234
OY 183 NTPLMSSRPNNPMVULPIFMHGKRNBSPTNSHARQFTPRNEFRAPRLKFOVSYKVAKE 242
Db 235 Q----KQKLEHPELQKQKHQKPS-BPELPLGKQOQESPE---PELPLGKQOQOQESPE 285
OY 243 DLWGTDSLDLWFGYTOQSHMOJFNCKNSRPRVHDYOPEITLTPVYSDLPMDGKVRMIGM 302
```

```
Db 286 -----PELQKQKQOQSHERPMDAGDQKQKQLH---KPELYLRKQOQYESP--DPE---LSL 333
OY 303 GAVHNSNGESAKLSRSMNRAYLMAGMEKNLTVMPRIMGRIFKGGSGSQPDNDPILDY 362
Db 334 GKQOHOEQOPELO-----LEEKQHKQPEPELHLGKQOQSHERPMDAED----- 377
OY 363 GYGDVRFYOLENKSNSISGTVRYNPRSGKALDLYVPLGKGISGYFOITQGYGOSL-- 420
Db 378 -----LEEKQKLG---BPELHLGKQOQOQIEREG-----YQGPKSLGOSLQK 416
OY 421 -----IDYNH 425
Db 417 EKASRQOQLDYSH 429
```

Search completed: November 30, 2001, 14:27:03
Job time: 550 sec

11 11 11

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:18:32 Search time 44.45 Seconds
(without alignments)
757,461 Million cell updates/sec

Title: US-09-787-083-4

Perfect score: 2360

Sequence: 1 MKVSLSTLTLSILPCFALIA.....YNHEATSPGVGLMDMMGL 442

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%
Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	815	34.5	382	2 E81195	phospholipase A1,
2	815	34.5	409	2 H81831	probable phospholip
3	373	15.8	286	2 B36971	outer membrane pho
4	366	15.5	329	2 D81279	phospholipase A1 (
5	358	15.2	289	2 A36971	outer membrane pho
6	357	15.1	289	1 P58CA1	phospholipase A1 (
7	321.5	13.6	289	2 E86069	outer membrane pho
8	246.5	10.4	355	2 C64582	phospholipase A1 p
9	240.5	10.2	355	2 H71930	probable phospholi
10	109.5	4.6	824	3 JC7532	cellulase (EC 3.2.
11	109.5	4.6	800	2 A29003	cellulase (EC 3.2.
12	109	4.6	822	2 JT0611	cellulase (EC 3.2.
13	106.5	4.5	602	1 TVKTRR	protein kinase (EC
14	101.5	4.3	783	2 JC5467	cellulase (EC 3.2.
15	98.5	4.1	1658	2 T42642	phospholipase A1 p
16	97.5	4.1	797	2 T46737	phospholipase A1 p
17	97.5	4.0	389	2 S76490	hypothetical prote
18	95.5	4.0	387	2 A35136	cellulase (EC 3.2.
19	94.5	4.0	601	2 T26062	hypothetical prote
20	94.5	4.0	655	2 D26061	hypothetical prote
21	94.5	4.0	1641	2 D82704	conserved hypotet
22	94.5	4.0	696	2 S55694	protein kinase (EC
23	94	4.0	765	2 T35719	chitinase - Strept
24	94	4.0	564	2 T40777	ferric reductase t
25	93.5	4.0	901	2 T20122	hypothetical prote
26	93.5	3.9	633	2 C81956	thiamin biosynthes
27	93	3.9	791	2 H96839	hypothetical prote
28	93	3.9	4273	2 C69679	polyketide synthas
29	93	3.9			

30	93	3.9	4307	2 T20721	hypothetical prote
31	92.5	3.9	857	1 A41369	S-receptor kinase
32	92.5	3.9	1807	2 T30940	vitellogenin - Pim
33	92.5	3.9	324	2 B69521	hypothetical prote
34	92	3.9	888	2 S50801	AMP deaminase homo
35	92	3.9	2971	2 T08026	hypothetical prote
36	92	3.9	5005	2 F82884	hypothetical prote
37	91.5	3.9	766	2 JC7355	peroxisome prolife
38	91	3.9	367	2 T24058	hypothetical prote
39	91	3.9	1078	2 T18352	protein P120 - Myc
40	91	3.9	1474	1 MAHU	alpha-2-macroglobu
41	90.5	3.8	386	2 A81328	probable periplasm
42	90.5	3.8	467	1 A49377	involucrin - mouse
43	90	3.8	841	2 JC5894	killer cell inhibi
44	90	3.8	1788	2 T29043	hypothetical prote
45	89.5	3.8	477	2 S23257	alpha-amyase (EC

ALIGNMENTS

Query Match	Score	815:	DB 2:	Length	382:
Best Local Similarity	44.7%				Pred. NO. 1.4e-59;
Matches	163;				Conservative 67; Mismatches 111; Indels 24; Gaps 8;
QY	87	INCSALNODIMRLACQDTLVHGETPAVY-----KTKRSIRLDETIMQTI-KGKPOVYVOE	140		
DB	33	LQCAALTDNVTIRLACQDTIFPAOLPSSAGQGESKAVLNTETVRSSLDGGEVIVVEK	92		
QY	141	TTDPFIFLWNGEKMTTKKDAKOLEYAAKQFTPLSFDLDRNN-TPLMSSRPHPMPVYL	199		
DB	93	GGDL-----PADSAGETADITYPLSLMDLDNDRLGLGVREHNPVWMLP	139		
QY	200	IFMHGKPNRSFNPFSHEAR-OFTPNERRAPLAKQVSKYKAAEDLMGTDSDLFGTQQ	258		
DB	140	LMVYNSPNYVAGSPTRGTVOEKRAETKIQVSEKSKIAEDLPKTRADLFGYTOR	199		
QY	259	SHMOIFN-GKNSRPFVHDVQPEIFLQPVYSDLPWGKVMIGMGVHNSGSAKLSR	317		
DB	200	SDMOIYNGKRSAPFRNTDVKPELFTQPVADLPFGKRLMLAGFVHNSGSRPESR	259		
QY	318	SWNRAYLMAGMEKMLNLTVMRPRINGRIKESGSDPDNDILDYGGDVFLQLENKS	377		
DB	260	SWNRITVYAMAGMEKMLNLTVMRPRINGRIKESGSDPDNDILDYGGDVFLQLENKS	317		
QY	378	NISGTVYRNPSGKGLQDLVYVPLGKIGISGFQIFQGYGOSLDLYNHEATSPGVGLMLN	437		
DB	318	NVYSVLRNPRTGYGALTAAYTFPIKGLKGVVGFHGYGSLIDYNNKONGIGIMFN	377		
QY	438	DMMGL 442			

OY 313 -----AKLSRSMNRAVYLMAGMEKNTVMPRTWGR 342
DB 424 VEDVYIENAGALKLISGLDASNDVSEGNWMAWRLSADGWMK 465

RESULT 13

JT0611
cellulase (EC 3.2.1.4), alkaline - Bacillus sp. (strain KSM-64)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus sp.
C:Date: 09-Oct-1992 #sequence_revision 09-Oct-1992 #text_change 10-Dec-1999
C:Accession: JT0611
R:Sunltomo, N.; Ozaki, K.; Kawai, S.; Ito, S.
Biosci. Biotechnol. Biochem. 56, 872-877, 1992
A:Title: Nucleotide sequence of the gene for an alkaline endoglucanase from an alkalophilic
A:Reference number: JT0611; MUID:92305459
A:Accession: JT0611
A:Molecule type: DNA
A:Residues: 1-822 <SUN>
A:Cross-references: GB:M04963; NID:9289264; PIDN:AAA73189.1; PID:9289266
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose
A:Pathway: cellulose degradation
C:Superfamily: Bacillus sp. alkaline cellulase; Thermotoga xylianse A amino-terminal ref
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:583-726/Domain: Thermotoga xylianse A amino-terminal repeat homology <TXA>

Query Match 4.6%; Score 109; DB 2; Length 822;
Best Local Similarity 19.7%; Pred. No. 1;

Matches 91; Conservative 51; Mismatches 150; Indels 170; Gaps 24;

OY 30 PVAEYDEVRSKND-----LGQDNELLIGVQASATQASTDTANPL-DEHEPELY-----TT 78
DB 25 PALAAEGNTREDNFKHLGNDNVRKPSRSEAGALQLEVDGQGTIVDQHEKIQLRGMSTH 84
OY 79 ALENTMLINCASL-----NQDIMRLACY-DTLVHGSTPAVITKRSIRDEIT----- 126
DB 85 GLQMPREILINDNAYKALANDMESNMIRLMAVYGENGYASNPPELLISRYIKGIDLAENDM 144
OY 127 -----WQT-----IKGKQVYVQETDP-----IFLMGN 150
DB 145 VYIVDMVHAPRDPDPVYAGAEDEFROIALYLPNNPHIYLANPSSNNNGAGCIRPN 204
OY 151 EKGMLTKKDAKQLEYAAKQFTPLSLFDLRN-----NTPLMSSRP-----HNPMY 196
DB 205 EEGMNAVVK-----EYADPTVEMLRDSGNADNIIIVGSPMNSQRPDLADNPIIDHHTWY 259
OY 197 VLPPIF--MHGKPNRS--PMTPSHEARQFTPNFRAPELKFQVSVKVKAAEDLMGT----- 247
DB 260 TVHETGSHASTESTYPPETPNSERGNWMSNTRYA-----LENGVAVFATE--WGTSQLANG 313
OY 248 -----DSDLWFGYTOQ-----SHMQIFNGKNS-----RPRRVH----- 275
DB 314 DGGPYFEADADVIEFLNENISMAMNSLTN--KNEVSGAFTPELKSNTSLDPPEDQYW 372
OY 276 -----DYOPEITLQPVYSDLPW--DGKVMIGMGAHVHNSGES- 312
DB 373 VPEELSLSGEYVRARIKGVNVEP--IDRTKYTKVLMDFNDTQKGFV-----NGDSP 423
OY 313 -----AKLSRSMNRAVYLMAGMEKNTVMPRTWGR 342
DB 424 VEDVYIENAGALKLISGLDASNDVSEGNWMAWRLSADGWMK 465

RESULT 14

TVTRRR
protein kinase (EC 2.7.1.37) raf - rat
N:Alternate names: kinase-related transforming protein raf; raf proto-oncogene protein-8
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 11-Jun-1999
C:Accession: B26126

R: Ishikawa, F.; Takaku, F.; Nagao, M.; Sugimura, T.
Mol. Cell. Biol. 7, 1226-1232, 1987
A:Title: Rat c-raf oncogene activation by a rearrangement that produces a fused prote
A:Reference number: A26126; MUID:87172791
A:Accession: B26126
A:Molecule type: mRNA
A:Residues: 1-602 <ISH>

A:Cross-references: GB:M15428; NID:9206546; PIDN:AAA42002.1; PID:9206547
C:Genetics:
A:Gene: raf
C:Superfamily: rat protein kinase raf; protein kinase homology
C:Keywords: ATP; autophosphorylation; phosphoprotein; phosphotransferase; serine/thre
F:301-567/Domain: protein kinase homology <KIN>
F:309-317/Region: protein kinase ATP-binding motif
F:329/Active site: Lys #status predicted
F:453/Binding site: phosphate (Ser) (covalent) (by autophosphorylation) #status predi

Query Match 4.5%; Score 106.5; DB 1; Length 602;
Best Local Similarity 19.0%; Pred. No. 1;

Matches 92; Conservative 74; Mismatches 172; Indels 145; Gaps 24;

OY 23 QOAVPNPVAFYDEVRSKNDLGQDNELLIGVQASATQASTDTANPLDEHEPELYTTALEN 82
DB 81 ELEKLNQYKVYTEKNKELETQADNRL-GIOSQTRAKEE-----LEAKRLDIT--N 131
OY 83 KTMLINCASLNDIMRLACQDTLVHGSTPAVITKRSIRDEITWQ--TIKGRPOVYQOE 140
DB 132 ERLSQEVEYLTEDVKRL--NEKLRESNT--TKGEIQTKDELQASDVYVKYREKRLQE 186
OY 141 -----TTDPTFLMGNEKG-----MLTKKQAKQLEYAAKQFTPLS 174
DB 187 KELLNQNSWLTETKTKTDELALGREGNEILEKCTLENKKEEDAIRSHSESASP 246
OY 175 LSFDDRNNTPLMSSRPNPMTVPLPIFMHGKPNRSPTPSHEARQFTPNFRAPELKFQV 234
DB 247 LS-SSPNMNSPTGQWQPKTP-----VPAQREKRAPGSTQCKNKRIRPGQDSSYWEI 298
OY 235 SVKVKAAEDLWGTSDSLWFGYTOQSHW-----QIFNGKNSRPFVHYDQPEI----- 281
DB 299 EASEVMLSTRIGSGS--FGTVYKGMHGDVAVKILKVVDPDPEDLOAFRNEVAVLKRT 355
OY 282 -----FLTO-----PYVSDL--PMDGKVRML-----GMGAVNH 307
DB 356 HVNILLPMGYMTKMDLAIVTQWCEGSSSLKHLVQETKFOMLDIDAROTAGMDYLHA 415
OY 308 SNGESAKLSRSMNRAVYLMAGMEK-----NLTVMPRTWGRIFKESGSGQ----- 351
DB 416 KNIIRHDKMS--NITFLHGLTVKIGDGLATVVKSRW-----SGSQQVEQPTGSVLM 466
OY 352 -----PDDNPDL--DYGGYGVRFYQLENKSNISGTVYINPRSGKGLQLDYVPL 402
DB 467 APEVIRMODNNPFSQSDVYSYIV--LYEL-----MTGELPYSHINNDDI--IFWV 515
OY 403 GKG 405
DB 516 GRG 518

RESULT 15

JC5467
cellulase (EC 3.2.1.4) - Bacillus sp.
N:Alternate names: Endo-1,4-beta-glucanase
C:Species: Bacillus sp.
C:Date: 07-Jul-1997 #sequence_revision 18-Jul-1997 #text_change 10-Dec-1999
C:Accession: JC5467
R: Miyatake, M.; Imada, K.
Biosci. Biotechnol. Biochem. 51, 362-364, 1997
A:Title: A gene encoding endo-1,4-beta-glucanase from Bacillus sp. 22-28.
A:Reference number: JC5467; MUID:97212038
A:Accession: JC5467
A:Molecule type: DNA
A:Residues: 1-783 <MIY>

A:Cross-references: DBJ:D85236

A:Experimental source: strain 22-28

C:Note: neither the complete nucleic acid sequence nor the complete translation are shown

C:Superfamily: Bacillus sp. alkaline cellulase; Thermotoga xylanase A amino-terminal repeat

C:Keywords: glycosidase; hydrolase

F:587-728/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA>

Query Match 4.38; Score 101.5; DB 2; Length 783;

Best Local Similarity 19.78; Pred. No. 3.9; Mismatches 148; Indels 169; Gaps 26;

Matches 91; Conservative 54; Mismatches 148; Indels 169; Gaps 26;

```

QY 30 PVAFVDEVRKND-----LGDNELLIGVQASASTDTANPL-DEHEPELY-----TT 78
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 25 PTALAEAGNTREDNFHLLNENYKRPSEAGALQKVEDGOMTLVDQHEKIQLRGMSTH 84
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 79 ALENKTMILNCSAL-----NODIMRLACYDTLVHGE-----TPAVIKTK-----RSI 120
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 85 GLQWFPEILNDNNAKALSDNDSNMIRLAMY---VGENGYATNPBELIKQRVIDIELAI 140
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 RLDETI---WQT-----IKGRQVYVYQETDP-----IF 146
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 141 ENDMYVIVDMHVAHPGDPDPVYAGAEDEFRDIAALYPNNPHTIYELANEPSSNNNGAG 200
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 147 LMGNEKMLTKDAKOLEYAAKQFTPLSLSFDDLRN---NTPLMSSRP-----H 192
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 201 IPNNEEGMKAVK-----EYADPIYEMLRDSGNADNDNIIYGSFPMSSQRPDLADNPINDH 255
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 193 NPMYVLEPIF--MHGKPNRS--PMTPSHEARQFTPNEFRAPBLKFQVSVKYKAAEDLWGT- 247
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 256 HTMYTVHFYTGSHAASTESYPPEPNSERGNVNSNTRYA---LENGVAVFATE--WGTS 309
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 248 -----DSDLWEGYTOQ-----SHWQIFNGKNS-----RPRV----- 274
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 310 QANGDGGPYDEADWIEFLNENNISWANNSLTN-KNEVSGAFTPFELGKSNATSLDGP 368
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 275 -----HDYQPEIFLTOPVYSGLPW---DGKVRMIGMGA----- 304
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 369 DQYWAPEELSLSGEYVARIKAKYEP---IDRTRYTKVLMDFNDCGKQGFVNSDSPNK 425
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 305 -----VHHSNGESAKLSSRNRAVILMAGMEWKNLTVMPRINGR 342
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 426 EAIEVENENG-TLRIISGLNVSNDLSGCFWANFRLSANGWGK 466
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: November 30, 2001, 14:18:34
Job time: 256 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:17:42 : Search time 36.79 Seconds
(without alignments)
270.358 Million cell updates/sec

Title: US-09-787-083-4

Perfect score: 2360

Sequence: 1 MKVSLSTLTSLIPCFAILA.....YNHEATSPGVGLMDNMGL 442

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/2/1aa/5a_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5b_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/5b_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfillsl.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101.5	4.3	1726	2	US-08-609-049A-30
2	101.5	4.3	1726	4	US-09-170-996-30
3	99.5	4.2	1658	2	US-08-609-049A-13
4	99.5	4.2	1658	1	US-09-170-996-13
5	89.5	3.8	857	1	US-07-717-331F-2
6	87.5	3.7	844	1	US-07-731-157A-6
7	87.5	3.7	844	2	US-08-541-780-6
8	87	3.7	355	1	US-07-946-497-5
9	87	3.7	355	2	US-08-483-322-5
10	87	3.7	355	1	US-08-478-882-5
11	87	3.7	503	1	US-07-946-497-2
12	87	3.7	503	1	US-08-483-322-2
13	87	3.7	503	2	US-08-478-882-2
14	86	3.6	1088	3	US-08-633-768A-1
15	84.5	3.6	1024	4	US-09-091-117-5
16	84	3.6	522	6	RE34606-6
17	84	3.6	1627	1	US-07-665-792E-9
18	83	3.5	494	1	US-08-275-488A-4
19	83	3.5	494	1	US-08-275-490-4
20	83	3.5	494	1	US-08-446-380-4
21	83	3.5	494	1	US-08-446-374-4
22	83	3.5	494	1	US-08-446-382-4
23	83	3.5	494	1	US-08-445-801-4
24	83	3.5	494	1	US-08-275-487-4
25	83	3.5	494	5	PCT-US95-08919-4
26	83	3.5	774	3	US-08-502-632-2
27	83	3.5	774	3	US-09-073-354-1

28	83	3.5	774	3	US-08-656-005A-1	Sequence 1, Appl
29	83	3.5	774	4	US-09-073-259-1	Sequence 1, Appl
30	83	3.5	774	4	US-09-363-095-1	Sequence 1, Appl
31	83	3.5	774	4	US-09-418-027-1	Sequence 1, Appl
32	82.5	3.5	657	4	US-09-306-593-2	Sequence 2, Appl
33	82	3.5	537	2	US-08-633-879C-2	Sequence 2, Appl
34	82	3.5	770	1	US-08-445-135-2	Sequence 2, Appl
35	81.5	3.5	663	1	US-08-441-139-7	Sequence 7, Appl
36	81.5	3.5	1022	3	US-08-772-270A-2	Sequence 2, Appl
37	81	3.4	649	2	US-08-871-266B-16	Sequence 16, Appl
38	81	3.4	649	2	US-09-018-864A-16	Sequence 16, Appl
39	81	3.4	649	3	US-08-871-267B-22	Sequence 22, Appl
40	81	3.4	649	4	US-09-618-419-22	Sequence 22, Appl
41	81	3.4	1121	1	US-07-789-915A-2	Sequence 2, Appl
42	81	3.4	1121	1	US-08-005-002C-2	Sequence 2, Appl
43	81	3.4	1121	1	US-08-487-203A-2	Sequence 2, Appl
44	80	3.4	334	2	US-08-359-850-4	Sequence 4, Appl
45	80	3.4	438	1	US-07-923-095-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-08-609-049A-30

Sequence 30, Application US/08609049A

Patent No. 5948664

GENERAL INFORMATION:

APPLICANT: Williams, Lewis T.

APPLICANT: Molz, Lisa

TITLE OF INVENTION: Chen, Yen-Wen

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/609,049A

FILING DATE: 29-FEB-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Dow, Karen B.

REGISTRATION NUMBER: 29,684

REFERENCE/DOCKET NUMBER: 2307K-063700US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-326-2400

TELEFAX: 415-326-2422

INFORMATION FOR SEQ ID NO: 30:

SEQUENCE CHARACTERISTICS:

LENGTH: 1726 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-609-049A-30

Query Match 4.3%; Score 101.5; DB 2; Length 1726;

Best Local Similarity 18.6%; Pred. No. 0.43;

Matches 98; Conservative 79; Mismatches 190; Indels 161; Gaps 21;

1 MKVSLSTLTSLIPCFAILAQQAAVNPVAFV-----EYRSNDLGGD 46

465 VKVSEIEGQLPVTFCIDVSVTEIITIMQALSVVHDLNVDVGSYILKVGCGEVLQN 524

NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 2307K-063700US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1658 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-609-049A-13

Query Match 4.2%; Score 99.5; DB 2; Length 1658;
Best Local Similarity 18.6%; Pred. No. 0.65;
Matches 98; Conservative 79; Mismatches 190; Indels 161; Gaps 21;

QY 1 MKVSL-----TLTSLPCFALLAQQAQVNPVAFVD-----EYRSKNDLGQD 46
DB 397 VKVSIETEGQLPYTFCDVSVTEIIMQALCWVHDDLQVDGSYLTKCGGEVLYQ 456
QY 47 NELLIGVQSASTPTANPLDEHEPELYTTALENTMLNCSALNODIMRLACYDTLV 106
DB 457 NHCLGSHHIONCKKMT-----EIKQLLTLSAMCONLARTEDD--- 497
QY 107 HGETPAVITKRSIRLDETIMQTKGRPOVVOYQETDPIF----- 146
DB 498 --EAP-----VDLKNYLQIEKPYKEVMIRHVEELDSYHYVELAQTENQHRAY 547
QY 147 -----LMGNEKMLTKKDAKOLEYAA-----KQFTPLSLFDDLRNTPLS 188
DB 548 DQVIAKARKISALDGVETPSVTEA-VKLLKRAVNLPRNSADVTSLSGSDTRKNSKGS 606
QY 189 SRPNPMTYLP-----IFMHGKPNR-----SPNTPSHARQFPPNEFRABELKQV 234
DB 607 LNPNPVOVSMDLHTLTAIYDLRLHANSSRCSTGCPGRSRIKAWATE-----QL 658
QY 235 SVKKAEDL--WGTDSDLMFGYTOQSHMOIFNGKN--SRPF--RVHDOPELFIQ-- 285
DB 659 QFTVYAAHGISSNMVSNVEKYLLCSLSH-----NGKDLFRQISKVGTAKYKPFYLLKMD 714
QY 286 -----PV-YSDLPMDGKVRMIGMGAVHHSNGESAKLRSNN-----RAYL 324
DB 715 ELITFPIQISQLPLESVLHTLTFGLVNLQSSGSPDSNQRKGRPEALGVSLTLPDFKRL 774
QY 325 MAGMEMKNTVMPRIWGRIFKEGSGSQDDNPDLIDYGYGDVDFLYOLEKNSISGTVR 384
DB 775 TCG-----TKLTYLW-----TSSHTNSIPGAIIPKKSVMERIVLQVDFPSPAFDIIT 821
QY 385 YNPRSGKALQLDVYVPLGKISGYFOIFQGYGSLIDYHNEATSGV 432
DB 822 TSPQIDRNIIQODKLETLESIDIKG-----KLDIHIRDSSFGL 859

RESULT 4
US-09-170-996-13
Sequence 13, Application US/09170996
Patent No. 6291220
GENERAL INFORMATION:
APPLICANT: Williams, Lewis T.
APPLICANT: Moitz, Lisa
APPLICANT: Chen, Yen-Wen
TITLE OF INVENTION: No. 6291220e1 PI 3-Kinase Polypeptides
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA

ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/170,996
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/609,049
FILING DATE: 29-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 2307K-063700US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2422
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1658 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-170-996-13

Query Match 4.2%; Score 99.5; DB 4; Length 1658;
Best Local Similarity 18.6%; Pred. No. 0.65;
Matches 98; Conservative 79; Mismatches 190; Indels 161; Gaps 21;

QY 1 MKVSL-----TLTSLPCFALLAQQAQVNPVAFVD-----EYRSKNDLGQD 46
DB 397 VKVSIETEGQLPYTFCDVSVTEIIMQALCWVHDDLQVDGSYLTKCGGEVLYQ 456
QY 47 NELLIGVQSASTPTANPLDEHEPELYTTALENTMLNCSALNODIMRLACYDTLV 106
DB 457 NHCLGSHHIONCKKMT-----EIKQLLTLSAMCONLARTEDD--- 497
QY 107 HGETPAVITKRSIRLDETIMQTKGRPOVVOYQETDPIF----- 146
DB 498 --EAP-----VDLKNYLQIEKPYKEVMIRHVEELDSYHYVELAQTENQHRAY 547
QY 147 -----LMGNEKMLTKKDAKOLEYAA-----KQFTPLSLFDDLRNTPLS 188
DB 548 DQVIAKARKISALDGVETPSVTEA-VKLLKRAVNLPRNSADVTSLSGSDTRKNSKGS 606
QY 189 SRPNPMTYLP-----IFMHGKPNR-----SPNTPSHARQFPPNEFRABELKQV 234
DB 607 LNPNPVOVSMDLHTLTAIYDLRLHANSSRCSTGCPGRSRIKAWATE-----QL 658
QY 235 SVKKAEDL--WGTDSDLMFGYTOQSHMOIFNGKN--SRPF--RVHDOPELFIQ-- 285
DB 659 QFTVYAAHGISSNMVSNVEKYLLCSLSH-----NGKDLFRQISKVGTAKYKPFYLLKMD 714
QY 286 -----PV-YSDLPMDGKVRMIGMGAVHHSNGESAKLRSNN-----RAYL 324
DB 715 ELITFPIQISQLPLESVLHTLTFGLVNLQSSGSPDSNQRKGRPEALGVSLTLPDFKRL 774
QY 325 MAGMEMKNTVMPRIWGRIFKEGSGSQDDNPDLIDYGYGDVDFLYOLEKNSISGTVR 384
DB 775 TCG-----TKLTYLW-----TSSHTNSIPGAIIPKKSVMERIVLQVDFPSPAFDIIT 821
QY 385 YNPRSGKALQLDVYVPLGKISGYFOIFQGYGSLIDYHNEATSGV 432
DB 822 TSPQIDRNIIQODKLETLESIDIKG-----KLDIHIRDSSFGL 859

RESULT 5
US-07-717-331F-2

```

; Sequence 2, Application US/07717331F
; Patent No. 5484905
; GENERAL INFORMATION:
; APPLICANT: June Nasrallah; Michael Nasrallah; and Joshua
; APPLICANT: Stein
; TITLE OF INVENTION: A Receptor Protein Kinase Gene
; TITLE OF INVENTION: Encoded At The Self-Incompatibility Locus
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yahwak & Associates
; STREET: 25 Skytop Drive
; CITY: Trumbull
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06611
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07717.331F
; FILING DATE: June 19th 1991
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: George M. Yahwak
; REGISTRATION NUMBER: 26,824
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203)268-1951
; TELEFAX: (203)268-1951
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 857 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-717-331F-2

Query Match          3.8%; Score 89.5; DB 1; Length 857;
Best Local Similarity 21.4%; Pred. No. 2.5;
Matches 76; Conservative 50; Mismatches 110; Indels 119; Gaps 23;

QY 105 LVHGERPAVIAKTRSRIRLEITMOTIKGRQVYVYQETTPIDPLMG---NEK---GMLTK 157
DB 24 LHPALSIYIYN---LSTESLTSSNKTLY---SPGSIIEVGFEFTNSRWYLGMYK 75
QY 158 K-DAKOLEYAAKQTPPLSLSPD-----LDRNNTPLMSSRPHNPMYVLPFMHGK 205
DB 76 KYSDRIYVAVARNRDNPLSLAIGTLKISGNNILVLDHSNKPVMWTN-----LTRGN 125
QY 206 PNRSP-----NTPSHEARQFT-----PNEFRAPDELKFOVSVKKAEDIM 245
DB 126 -RSRPAVIELLANGNEVMDSSNNDASEYIMOSFDYPTDTLLPEMKLGYNLKT----- 177
QY 246 GTRSDIMEGYTQOSHQIENGKNSRFRVHDYQ-----PEIFLQ---PVYSDLPMDCK 296
DB 178 GLNRFL-----TSKRSDDPSSGNF---SYKLETSQSLPEFYLSKRNPMHRSQWNG- 226
QY 297 VEMIGGAVH-----HSNGESAKLSRSMNRAY-----LMAGMEKNTLVMP--RI 339
DB 227 IFFSGIPEDQKLSYMYVNIENNEEVAAYFRMTNNSFYRLTLISISGQRLTLWPSIRI 286
QY 340 WQRIRKESGSGSPD---NPDILDYGYGDVRFELYQLENKSNISGTVR-VNPRS 389
DB 287 WNRFESSPVDRQCDTYIMCGP-----YAYCDV-----NTSPVCNCIGQFNPRN 329

```

```

; APPLICANT: Quax, Wilhelmus J.
; APPLICANT: Misset, Onno
; APPLICANT: Van der Laan, Jan M.
; APPLICANT: Lenting, Herman B.M.
; TITLE OF INVENTION: Mutated beta-lactam acylase genes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GOMARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (Epo)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/731,157A
; FILING DATE: 19910509
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90200962
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: RAE-VENTER PH.D., BARBARA
; REGISTRATION NUMBER: 32,750
; REFERENCE/DOCKET NUMBER: GBRO-027/00US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-494-7622
; TELEFAX: 415-857-0663
; TELEX: 380816 COOLEY PA
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 844 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Kluyvera citrophila
; STRAIN: ATCC 21285
; US-07-731-157A-6

Query Match          3.7%; Score 87.5; DB 1; Length 844;
Best Local Similarity 21.2%; Pred. No. 3.9;
Matches 92; Conservative 46; Mismatches 152; Indels 143; Gaps 25;

QY 31 VAFVDEVRSK--NDIGQD--NEL--LIGVOSATQASSTDTANPLDEHEPELYTTAL----- 80
DB 182 LALLTAVKKDKYNDDEGMAVFNQKLWNPASPTTIARESSYPLKFDLQNTOTALLVPR 241
QY 81 -ENKTMILNCSALNDIMLACYDYLIVHETPAVIAKTRKISILDETIWQITGKQPVVYQ 139
DB 242 YQAPAMLDLRAKGTGALLAV-----TAIKNRTTIA-----OFANGANGLACY 286
QY 140 ETTDPIFLMGNEKGMILTKDAKOLEYAAKQTPPLSLSPDLDRNNTPLMSSRPHNPMYVLP 199
DB 287 PTTSMNVAVIGKKA---QDAKAIWNGPQFG-----W---YAPAYTVG 323
QY 200 IFMHGKP-NRSPNTSHEARQFTPNEFRAPDELKFOVSVKKAEDIMGT-----DSDIM 252
DB 324 IGIHAGYDVNTGNTP-----FAYPGLVFGHNGTIS---WGSTAGFGDQVDIF 367
QY 253 F-----GYTQOSHQIENGKNSRFRVHDYQPEIFL-----TQPVYS 289
DB 368 AEKLSAEKPGYTOHNGEWKMLSRKETTIAVKDQGPETFTVMFTLGGNVIKTDTROTATA 427
QY 290 DL-PMDGKVRMIGAVHHSNGESAKLSRSMNRAVILMAGMEKNTLV-----VMPRIW-- 340
DB 428 KARAWAGK-----EVA5L-LAW--THOMKAKKNPENTQQAQKALITINWY 470

```


Query Match	3.7%;	Score 87;	DB 1;	Length 355;
Best Local Similarity	19.3%;	Pred. NO. 1.1;		
Matches	59;	Conservative	36;	Mismatches 97;
				Indels 114;
				Gaps 16

QY 45 QDNELLIG--VQSATQASSTDTA-NPLDEHEPELYTTALENKTMLNCSALNODIMRLAC 101

```

Db      7  RDAFFIGSTLATSTESNTNPTGKWPNEENEDE-----TDKYPNFSGSGIDDD
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

102 YDTLVHGETPAV-IKTKRSIRLDETIWQTIKGPQVVYQETTDPIFLMGNEKGMTKKDA 160

```

Db      57 FISSIATTPWVSAHTKON--QERTQWNPRIHSNPEVLLQTTT----- 96

```

Q: 161 FOR EVALUATION OF ANTI-DOPING MEASURES 010

00 101 A0LEIAA0FIPLESFDDDRNNIPLE---WSSRPHNPMYVLPLEMHGKPNRSPNIPSSHE 210

```
Db 97 -----RMT-----DIDRNSTAHGENWTOEPQPPF-----NNHEYQ 127
```

00 317 ABOETPNEFBABETKEOVSVKVKAABEDIW-----CTDCDNTWECVTOOOSHWOTENC 366

21	NAME	LAST	FIRST	MIDDLE	INITIAL	DATE	TIME	200
22	:	:	:	:	:	:	:	

Db 128 DEETPH-----ATSTWADPNSTTEAATQKEKWF-----ENEMQ---G 164

OV 267 KN-----SRPFRVHDYOPETIETOPVYSDLPWDGKVRMIG--MGAVHHSNGES 312

D_b 165 KNPPTSEDSHVTEGTTASAHNNHPQSRMTTQSQEDVSWTFDDPDPHMGQGQHOTESKD 224
QY 313 AKLSRS 318
D_b 225 TGSSTHS 230

```

;      TYPE:  amino acid
;      TOPOLOGY:  linear
;      IMMEDIATE SOURCE:
;      CLONE:  rat protein
US-08-483-322-5

```

Query Match	3.78;	Score 87;	DB 1;	Length 355;
-------------	-------	-----------	-------	-------------

Best Local Similarity	19.38;	Pred. No. 1.1;
Matches	59: Conservative	36: Mismatches
		97: Indels
		114: Gaps

MEMBERS	20	CONSTRUCTIVE	20	MEMBERS	21	INETS	44	34
MEMBERS	20	CONSTRUCTIVE	20	MEMBERS	21	INETS	44	34

45 QDNELLIG--VQATQASTDTA-NPLDEHEPELYTTALENKTMLINCSALNQDIMRIAC 101

```
7 RDAFFIGSTLATSTESNTPTGKNEENEDE-----TDKYPNFSGSGIDPD-----ED 56
```

102 YDVLVHGETPAV-IKTKRSIRLDETIWQTIKKGKPVVYQETTDPIFLMGNEKGMLTKKDA 160

```

57 FISSIATTPWWSAHTKON--QERTOWNPIHSNPEVLLQTTT----- 96

```

[illegible]

```
101 KQLEYAAQTTPESLFDLDKRNIFPL---WSSRPHNPMIVLPLEMHGKPNKRSPNIPSEI 216
```

97 -----RMT-----DIDRNSTSAHGENWTQEPQPF-----NNHEYQ 127

217 ABOETDNEERAPET.KEOVSVKVKA AFDI W-----CTDSDI WEGYTONSHWOTENG 366

[illegible]

Dd 128 DEETPH-----ATSTTMADPNSTTEAATQEKWPF-----ENEMQ---G 164

Qy 267 KN-----SRPRVHDYDPELFLTPRYSIDLPMDCGYKRMIG--MCAVHHNSGES 312

Db 165 KNPSPSEDSHVEGTGTAAAHNNHPSQRMTTQSQEDVSMTDFDPRIshPMGCQHQTESKD 224

```

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids

```


FILING DATE: 19921109

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

...

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:16:58 ; Search time 72.04 Seconds
(without alignments)
454,475 Million cell updates/sec

Title: US-09-787-083-4

Perfect score: 2360

Sequence: 1 MKVSLSTFTLSILPCFATLA.....YNHEATSRGVLMLNDWGL 442

Scoring table:

BLOSUM62

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A.Geneseq-1101:*

1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT:*
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT:*
19: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT:*
20: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT:*
21: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT:*
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2360	100.0	442	21	BAAB034 amino acid
2	2339	99.1	442	21	BAAB034 amino acid
3	2330	98.7	442	21	BAAB034 amino acid
4	2330	98.7	442	21	BAAB034 amino acid
5	822	34.8	370	21	BAAB034 amino acid
6	815	34.5	370	21	BAAB034 amino acid
7	815	34.5	370	21	BAAB034 amino acid
8	815	34.5	370	21	BAAB034 amino acid
9	812	34.4	375	21	BAAB034 amino acid
10	246.5	10.4	355	19	AAV70628
11	239.5	10.1	356	19	AAV10960

12	156.5	6.6	253	18	AAW20760
13	109.5	4.6	824	21	AAAB23180
14	109	4.6	800	8	AAV70420
15	109	4.6	822	13	AAAB26021
16	107	4.5	157	18	AAW20538
17	106	4.5	798	21	AAAB40925
18	106	4.5	798	21	AAV90225
19	106	4.5	798	21	AAAB3944
20	105.5	4.5	537	22	AAAG1000
21	102.5	4.3	166	19	AAW70991
22	98.5	4.2	467	20	AAV13378
23	98.5	4.2	467	21	AAV94860
24	98.5	4.2	467	22	AAAB8401
25	98.5	4.2	467	22	AAAB8026
26	98.5	4.2	1726	18	AAAB38756
27	97	4.1	761	20	AAW99084
28	95.5	4.0	502	22	AAAG90031
29	95.5	4.0	516	22	AAAB78876
30	95	4.0	522	22	AAAB95616
31	94	4.0	372	16	AAAB69607
32	93.5	4.0	467	19	AAW75057
33	92	3.9	888	22	AAAG70751
34	92	3.9	1484	12	AAAB11749
35	91.5	3.9	857	13	AAAB29814
36	91	3.9	600	22	AAAG75092
37	91	3.9	1227	22	AAAB81501
38	91	3.9	1474	21	AAV97157
39	91	3.9	1474	21	AAAB50673
40	90	3.8	522	22	AAAB95513
41	90	3.8	564	21	AAAB10457
42	90	3.8	564	21	AAAB1952
43	90	3.8	790	22	AAAB93045
44	90	3.8	841	20	AAV24318
45	90	3.8	1627	16	AAAB67538

ALIGNMENTS

RESULT 1

AAV85269 standard; Protein: 442 AA.

AC AAV85269;

XX 29-JUN-2000 (first entry)

XX BASB034 amino acid sequence #2.

XX DE

XX Moraxella catarrhalis infection; BASB034; diagnosis; straging;

KW vaccine; bacteriostatic; treatment; prevention; otitis media; pneumonia;

KW sinusitis; nosocomial infection; invasive disease; chronic otitis media;

KW hearing loss; antibacterial drug.

XX Moraxella catarrhalis.

XX WO200015802-A1.

XX PD 23-MAR-2000.

XX PF 14-SEP-1999; 99MO-EP06781.

XX PR 14-SEP-1998; 98GB-0020002.

XX (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Ruelle J;

XX WPI; 2000-271440/23.

XX N-PSDB; AAA10701.

XX Novel BASB034 polynucleotides and polypeptides from Moraxella

PT catarrhalis used to prepare vaccines against bacterial infections


```

PR 14-SEP-1998; 98GB-0020002.
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Ruelle J;
XX WPI: 2000-271440/23.
DR N-PSDB: AAA10702.
XX
PT Novel BASB034 polynucleotides and polypeptides from Moraxella
PT catarrhalis used to prepare vaccines against bacterial infections
XX
PS Claim 3; Page 68; 106pp; English.
XX
CC This sequence represents a Moraxella catarrhalis BASB034 polypeptide from
CC strain Mc2913. The invention relates to BASB034 polypeptides from
CC M. catarrhalis strains Mc2913, Mc2908, Mc2913 and Mc2969. The BASB034
CC polynucleotides and polypeptides may be employed as research reagents and
CC material for the discovery of treatments and diagnostics for diseases,
CC particularly human diseases. They are particularly used to diagnose and
CC treat M. catarrhalis infections. They can be used for diagnosis of
CC disease, staging of disease, or determining response of an infectious
CC organism to drugs. The polynucleotides may be used as a source for
CC hybridization probes, and for screening of genetic mutations, serotype,
CC organism or strain identification, identification of mutations in BASB034
CC sequences, and as components of arrays which are useful for diagnostic
CC and prognostic purposes. The polypeptides can be used to produce
CC antibodies. The polypeptides can also be used in vaccine formulations,
CC and to identify agonists and antagonists. The polypeptides, antibodies,
CC agonists and antagonists (which are bacteriostatic) are used for the
CC treatment and prevention of diseases such as otitis media in infants and
CC children, pneumonia in elders, sinusitis, nosocomial infections and
CC invasive diseases, and chronic otitis media with hearing loss. The
CC polypeptides, agonists and antagonists are also used for screening of
CC antibacterial drugs. The BASB034 products of the invention can be used
CC screen for new antibacterial compounds that may target resistant
CC bacteria.
XX
SQ Sequence 442 AA:

```

```

Query Match 98.7%; Score 2330; DB 21; Length 442;
Best Local Similarity 98.9%; Pred. No. 2,2e-216;
Matches 437; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 1 MKVSLSTLTSLPCFALIAIQQAVPNPAFVDEVRKNDLGDNDNELLVGQATQSA 60
DB 1 mkvslstltslscfalfalqakavnpvafvdevrsendlgqdnelpdvgsatgsa 60
OY 61 SUTDTANPIDEHEPELTYTALEKTKLINSALNODIMRLACVDTLVHGTPRAVITKRSI 120
DB 61 stctanpidehepeltytalenkmlncsalngdimrlacvdtlvhgtpaviktrsi 120
OY 121 RLDETIMOTIKKPOVVOETDPIFLMGNEKMGMLTKKDAKOLEYAAKOFPLSIFD 180
DB 121 rldetlmotikkgpvovvgetdpriflmgnekgmltkkdakleyaakofplslsfld 180
OY 181 RNNPFLWSSRRPHNPMYVLPPIEMHGKPNNSPNTPSHEAKOFPPNEFRAPLKEQVSVKYA 240
DB 181 rnnpflwssrrphnpyvlppliemhgkpnnsptpsheakofppnefrapelkqysvkvka 240
OY 241 AADLWCTSDLMFGYTOOSHQOIFNGKNSRRPRVHDYQPEITLVQPVYSDDLWDDKVRKI 300
DB 241 aedlwctsdldlfgyltqshwqifngknsrprvhdypetlvpvysddlwdkvrki 300
OY 301 GNGAVHNSGESAKLSRSNRAYLMAEMKMLTVPRIWGRIFKEGSGSQDDNDPID 360
DB 301 gngavhnsgeasaklsrsnraylmagmewkmltvmprlwgriifegsgsqddndpid 360
OY 361 YYGVDVRFVLVYLENKSNIIGTVRYNPNRSGKALQDLVYVPLGKISGYSFOITQGYGSL 420
DB 361 yygvdvrfvllylenknsniigstvrynpnrsgkalqdlvyvplgkisyfqiifggygs 420

```

```

OY 421 IDYNHEATSPGVGLMLNDMGL 442
DB 421 idynheatsfgvglmndmgl 442

```

RESULT 5

```
AAV75156
ID AAV75156 standard; Protein; 370 AA.
XX
XX AAV75156;
XX
XX 21-MAR-2000 (first entry)
XX
XX Neisseria gonorrhoeae ORF 582 protein sequence SEQ ID NO:1786.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX
XX antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
XX
XX antibacterial; gene therapy.
XX
XX Neisseria gonorrhoeae.
XX
XX W0957280-A2.
XX
XX 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
XX
XX 31-JUL-1998; 98US-0094869.
XX
XX 02-SEP-1998; 98US-0098994.
XX
XX 02-SEP-1998; 98US-0099062.
XX
XX 09-OCT-1998; 98US-0103749.
XX
XX 09-OCT-1998; 98US-0103794.
XX
XX 09-OCT-1998; 98US-0103796.
XX
XX 25-FEB-1999; 99US-0121528.
XX
XX
XX (CHIR ) CHIRON CORP.
XX
XX (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX
XX Petersen J, Piazza M, Rappuoli R, Ratcl G, Scalato E, Scarselli M;
XX
XX Tettelin H, Venter JC;
XX
XX WPI: 2000-062150/05.
XX
XX N-PSDB: AA253918.
XX
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX
XX vaccines and diagnostics
XX
XX Claim 2; Page 903; 1453pp; English.
XX
XX AA254535 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941
XX
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX
XX CC and primers used in the exemplification of the present invention. The
XX
XX PCR primers used in the exemplification of the present invention. The
XX
XX polypeptides, the polynucleotides, antibodies and compositions of
XX
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX
XX immunogenic compositions. The polypeptides can be used in the
XX
XX manufacture of medicaments for treating or preventing infection due to
XX
XX Neisseria bacteria (e.g. meningitis and septicaemia), to detect the
XX
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX
XX be used to screen for agonists or antagonists, which may themselves
XX
XX have use as antibacterial agents. The polynucleotides of the invention
XX
XX may also be used in gene therapy protocols.
XX
XX
XX Sequence 370 AA:

```

```

Query Match 34.8%; Score 822; DB 21; Length 370;
Best Local Similarity 44.7%; Pred. No. 8.5e-71;
Matches 163; Conservative 68; Mismatches 110; Indels 24; Gaps 8;
OY 87 INCSALNDIMRLACVDTLVHGTPRAVITKRSIRDETIMOTI-KKKPQVVOE 140

```

```

Db      21  lqcaaltchvtrlacysdrifaaqlpsasaggegskevavlnltetvssldkgeavivvek 80
QY      141  TTPPIFLMGNEKMLTKKDAKOLEYAAKQFTPLSLSPDLDRNN-TPLMSSRPHPMVLVP 199
Db      81  ggda1-----padsagetadlycpislmdyldkndrgllyvrehmpmlymp 127
QY      200  IFMHGKPNRSPNTPSHEAR-QFTPNFEFRAPELKFQVSVKVAABDLMGTSDDLMEFGYTOQ 258
Db      128  fwnynspnyapsaprtgtvtgqekfgrkretklqvsfkkskaenlftktradiwfygtqr 167
QY      259  SHWOIEN-GKNSRPFVRYHDYQPEITFLQPVYSDLPMWDGKVRMTGMAVHNSGESAKLSR 317
Db      188  sdwqilyngfksapfrntdykpeifltqpvkadlprfggrlrmlyagfvtvhsngsqsrpre 247
QY      318  SWNRAYLMAEMENKNTLVMRIRMGRIKFGSGSQPDNDPDLIDYGYGDAVRFYLQLENKS 377
Db      248  swnrilyamagmewgklvtrvwvratdq-sgdk-ndnpdladymgvgdvklqyrlndrq 305
QY      378  NISGTVRYNPRSGKALQDLVYVPLGKISGYFOIFOGYGOSLIDYNHGATSFGVGLMLN 437
Db      306  nysvsvlrynpktygaleaaytfpikgklkvvgvrgfngygeslidyhnhkngigilmfn 365
QY      438  DWMGL 442
Db      366  dwdgl 370

RESULT  6
AA75157
ID  AA75157 standard; Protein; 370 AA.
XX
AC  AA75157;
XX
DT  21-MAR-2000 (first entry)
XX
DE  Neisseria meningitidis ORF 582 protein sequence SEQ ID NO:1788.
XX
KW  Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM  antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW  antibacterial; gene therapy.
XX
OS  Neisseria meningitidis.
XX
PN  MO9957280-A2.
XX
PD  11-NOV-1999.
XX
PF  30-APR-1999; 99WO-US09346.
XX
PR  01-MAY-1998; 98US-0083758.
PR  31-JUL-1998; 98US-0094869.
PR  02-SEP-1998; 98US-0098994.
PR  02-SEP-1998; 98US-0099062.
PR  09-OCT-1998; 98US-0103749.
PR  09-OCT-1998; 98US-0103794.
PR  09-OCT-1998; 98US-0103796.
PR  25-FEB-1999; 99US-0121528.
XX
XX  (CHIR ) CHIRON CORP.
PA  (GENO-) INST GENOMIC RES.
XX
PI  Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
PI  Petersen J, Pizzo M, Rappelli R, Ratti G, Scalato E, Scarselli M;
PI  Tettelin H, Venter JC;
XX
DR  MPI; 2000-062150/05.
XX
DR  N-PSDB; AA253919.
XX
PT  Novel Neisserial polypeptides predicted to be useful antigens for
XX  vaccines and diagnostics
XX
PS  Claim 2; Page 904; 1453pp; English.
```

```

XX
CC  AA253015 to AA254536, AA254577 to AA254615, and AA774253 to AA759941
CC  represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC  and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
CC  PCR primers used in the exemplification of the present invention. The
CC  polypeptides, the polynucleotides, antibodies and compositions of
CC  the invention can be used as vaccines, as diagnostic reagents, and as
CC  immunogenic compositions. The polypeptides can be used in the
CC  manufacture of medicaments for treating or preventing infection due to
CC  Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC  presence of Neisseria bacteria, or to raise antibodies. They may also
CC  be used to screen for agonists or antagonists, which may themselves
CC  have use as antibacterial agents. The polynucleotides of the invention
CC  may also be used in gene therapy protocols.
XX
SO  Sequence 370 AA;
XX
Query Match 34.5%; Score 815; DB 21; Length 370;
Best Local Similarity 44.7%; Pred. No. 4.1e-70;
Matches 163; Conservative 67; Mismatches 111; Indels 24; Gaps 8;
QY  87  INCSALMODIMRLACYTTLVHGETPAVI-----KTKRSIRLDETITQGT-KGRQVYVYQE 140
Db      21  lqcaaltchvtrlacysdrifaaqlpsasaggegskevavlnltetvssldkgeavivvek 80
QY  141  TTPPIFLMGNEKMLTKKDAKOLEYAAKQFTPLSLSPDLDRNN-TPLMSSRPHPMVLVP 199
Db      81  ggda1-----padsagetadlycpislmdyldkndrgllyvrehmpmlymp 127
QY  200  IFMHGKPNRSPNTPSHEAR-QFTPNFEFRAPELKFQVSVKVAABDLMGTSDDLMEFGYTOQ 258
Db      128  lwnynspnyapsaprtgtvtgqekfgrkretklqvsfkkskaenlftktradiwfygtqr 187
QY  259  SHWOIEN-GKNSRPFVRYHDYQPEITFLQPVYSDLPMWDGKVRMTGMAVHNSGESAKLSR 317
Db      188  sdwqilyngfksapfrntdykpeifltqpvkadlprfggrlrmlyagfvtvhsngsqsrpre 247
QY  318  SWNRAYLMAEMENKNTLVMRIRMGRIKFGSGSQPDNDPDLIDYGYGDAVRFYLQLENKS 377
Db      248  swnrilyamagmewgklvtrvwvratdq-sgdk-ndnpdladymgvgdvklqyrlndrq 305
QY  378  NISGTVRYNPRSGKALQDLVYVPLGKISGYFOIFOGYGOSLIDYNHGATSFGVGLMLN 437
Db      306  nysvsvlrynpktygaleaaytfpikgklkvvgvrgfngygeslidyhnhkngigilmfn 365
QY  438  DWMGL 442
Db      366  dldgl 370

RESULT  7
AA75158
ID  AA75158 standard; Protein; 370 AA.
XX
AC  AA75158;
XX
DT  21-MAR-2000 (first entry)
XX
DE  Neisseria meningitidis ORF 582 protein sequence SEQ ID NO:1790.
XX
KW  Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM  antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW  antibacterial; gene therapy.
XX
OS  Neisseria meningitidis.
XX
PN  MO9957280-A2.
XX
PD  11-NOV-1999.
XX
PF  30-APR-1999; 99WO-US09346.
XX
```


[illegible]

DB	193	sdwqlyngqrktsaprlndykrpeltcpvkadlpfagrlmlgagfvnqnsgrprear	252
QY	318	SNNNRAVTLAAGMEWKNLTVPRIMGRIFKFGSSQDPDNDIIDYGGCVRFLLYOLENKS	377
DB	253	swvrllyamagmewgkltvprvwvrafdq-sgdk-ndnpdiadymgavdklqyrlndrq	310
QY	378	NISGVRYRVPRSGKGLDLDYVYPIGKGISGYFQIFQGYGOSLIDVNHFATSFGVGLMNL	437
DB	311	nuysvrlftrpkrctgysaleaaylrlpkrklykgyvrgfhyggeslidyhmkqnglglmltn	370
QY	438	DMWGL 442	
DB	371	dldgl 375	
RESULT 10			
ID	AAW98871		
XX	AAW98871	standard; Protein; 355 AA.	
XX	AAW98871;		
AC	AAW98871;		
DT	31-MAR-1999	(first entry)	
XX			
DE	H. pylori	GHPO 1723 protein.	
XX			
KW	GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;		
KM	peptic ulcer disease.		
XX			
OS	Helicobacter pylori.		
XX			
PN	WO9843478-A1.		
PD			
XX	08-OCT-1998.		
PF	01-APR-1998;	98WO-US06371.	
XX			
PR	29-JUL-1997;	97US-0902615.	
PR	01-APR-1997;	97US-0833457.	
PR	24-JUN-1997;	97US-0801227.	
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
PA	(INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.		
XX			
PI	Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;		
XX			
DR	WPI: 1998-542293/46.		
DR	N-PSDB: AAX14590.		
XX			
PT	New isolated Helicobacter polynucleotides - used to develop products		
PT	for the diagnosis, prevention and treatment of Helicobacter		
PT	infections and gastrointestinal diseases		
XX			
PS	Claim 8; Page 1976-1977; 2054pp; English.		
XX			
CC	This sequence represents a Helicobacter pylori GHPO protein of the		
CC	invention. The polypeptides can be used for preventing or treating		
CC	Helicobacter infections, and gastroduodenal diseases associated with		
CC	these infections, including acute, chronic, and atrophic gastritis, and		
CC	peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be		
CC	used for the production of antibodies. The products can also be used for		
CC	detection and diagnosis.		
XX			
XX			
SO	Sequence	355 AA;	
Query Match			
Best Local Similarity 25.1%; Score 246.5; DB 19; Length 355;			
Matches 82; Conservative 41; Mismatches 117; Indels 87; Gaps 11;			
QY	157	KKDAQOLEAAAOQFPLSLSPDLDDNNRNTPLMSRRPHNPYVLPDIFMHCGRNPSPTPSHE	216
DB	69	kyllmmddylgtvflpbfyhsf-----fpdlgtwyhprlnp-----	102

```

QY 217 AROETPRNEFARPELKFOVSVYKAAEDLMGDDSLMEGYQOQSHWOJENKKNRPPRYHD 276
Db 103 ---ygrnef-----kfgstfvrpvrfnhlwckagtylalyaqtdwtdqlymnpqgsapmrmm 154
QY 277 YQPEIFLTQPYVSDLPWDGKV--RMIGGAVHNHNSG--ESAKLSRSKNRAYILMAGHEWKN 332
Db 155 fmpelilyuypr-nfkprfgsktgmfseilwlgwqhlnsvgagcqyprfk----egnpeng 209
QY 333 LTVMPR-----WGRIFEGSGSR-----DDNPD 357
Db 210 fprgqrvlkkdyngskdvrtwggcstsvsagqrprvflrlwckgglklymawprryvdqsnpr 269
QY 358 TLDYVYGCDVRFY-----OLENKSNISGTVAYNRPNSKGALDLDYVPLKGISGYE 410
Db 270 lldymygnakldyrtgrhnheldlylalficqwyd--rwhgafrlgytuyrnpfvgly 327
QY 411 QIFQGYGSLDLYNHNEATSPVGLMLN 437
Db 328 qwfngygdglyeydvfnrclgvglrln 354

```

```

RESULT 11
AAV10960
ID AAV10960 standard; Protein; 356 AA

```

Query Match	10.1%;	Score 239.5;	DB 19;	Length 356;
Best Local Similarity	24.8%;	Pred. No. 1.5e-14;		
Matches	82;	Conservative 43;	Mismatches 111;	Indels 95; Gaps 12

```

0Y 157 KKAQQLUAAKQKQFRLSLSDLRNNRPLMSSNRPNRMVULFEMHGKRPNSPTFSHE 216
Db 70 kYUmmmdyUlbYfLfrfshf-----tPlfwYhnpInP----- 103
0Y 217 AROFNPNEFRARBELFQOVSUVUKAAEDLMGDSOLMEGTYOOSMOQIFNGKNSRPRFVND 276
Db 104 ---ygrnef-----kfgslstfvrfrfhllksgcllyayqctmwfqjyndprqesamtmIn 155
0Y 277 YQELFELTQPVYUSSLRPMWDGKY--RMIGGAVHNHNSG--ESAKLSRSMNRA----- 322
Db 156 fmprellyuyrL-nfKfrfgskLgnfseklayqrlhslngvgagqcyqrlnkegnpenqfprq 214
0Y 323 -----YLMAGHEMK--NLTVMPRIWGRIFKEGSGSDPDD- 354
Db 215 pvlvkdngqrkdvtwgscsxvxxgnxlcsfvlwaeqgklkImvaywpyv-----pydq 266
0Y 355 -NPDLIDLUGYGQDVAFL-----QLENKSNIISGYTRKYNRBSKGAOLDYUPLRLKGI 406
Db 267 snpqldlyumygnakldyrtgrhlfelqldyqlficywtgd--twhgaftrlyuytgrInfv 324
0Y 407 SGQFIQSGQSLDLUDNHMTRSPGVGMLN 437
Db 325 glYaqwfmngydgdlYevdfnslayvqdlrIn 355

```

RESULT 12
AAW20760
ID AAW20760 standard; protein; 253 AA

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:10:38 ; Search time 72.04 Seconds
(without alignments)
454.475 Million cell updates/sec

Title: US-09-787-083-2

Perfect score: 2360

Sequence: 1 MKVSLSTLTSLTSLSCFAILA.....YNHEATSFQVGLMINDMGL 442

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.GeneSeq_1101:*

- 1: /SIDSB/gcgdata/geneSeq/geneSeq/AA1980.DAT:*
- 2: /SIDSB/gcgdata/geneSeq/geneSeq/AA1981.DAT:*
- 3: /SIDSB/gcgdata/geneSeq/geneSeq/AA1982.DAT:*
- 4: /SIDSB/gcgdata/geneSeq/geneSeq/AA1983.DAT:*
- 5: /SIDSB/gcgdata/geneSeq/geneSeq/AA1984.DAT:*
- 6: /SIDSB/gcgdata/geneSeq/geneSeq/AA1985.DAT:*
- 7: /SIDSB/gcgdata/geneSeq/geneSeq/AA1986.DAT:*
- 8: /SIDSB/gcgdata/geneSeq/geneSeq/AA1987.DAT:*
- 9: /SIDSB/gcgdata/geneSeq/geneSeq/AA1988.DAT:*
- 10: /SIDSB/gcgdata/geneSeq/geneSeq/AA1989.DAT:*
- 11: /SIDSB/gcgdata/geneSeq/geneSeq/AA1990.DAT:*
- 12: /SIDSB/gcgdata/geneSeq/geneSeq/AA1991.DAT:*
- 13: /SIDSB/gcgdata/geneSeq/geneSeq/AA1992.DAT:*
- 14: /SIDSB/gcgdata/geneSeq/geneSeq/AA1993.DAT:*
- 15: /SIDSB/gcgdata/geneSeq/geneSeq/AA1994.DAT:*
- 16: /SIDSB/gcgdata/geneSeq/geneSeq/AA1995.DAT:*
- 17: /SIDSB/gcgdata/geneSeq/geneSeq/AA1996.DAT:*
- 18: /SIDSB/gcgdata/geneSeq/geneSeq/AA1997.DAT:*
- 19: /SIDSB/gcgdata/geneSeq/geneSeq/AA1998.DAT:*
- 20: /SIDSB/gcgdata/geneSeq/geneSeq/AA1999.DAT:*
- 21: /SIDSB/gcgdata/geneSeq/geneSeq/AA2000.DAT:*
- 22: /SIDSB/gcgdata/geneSeq/geneSeq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2360	100.0	442	21	BAASB034 amino acid
2	2354	99.7	442	21	BAASB034 amino acid
3	2352	99.7	442	21	BAASB034 amino acid
4	2330	98.7	442	21	BAASB034 amino acid
5	821	34.8	370	21	BAASB034 amino acid
6	814	34.5	370	21	BAASB034 amino acid
7	814	34.5	370	21	BAASB034 amino acid
8	814	34.5	370	21	BAASB034 amino acid
9	811	34.4	374	21	BAASB034 amino acid
10	246.5	10.4	355	19	AAW98871
11	239.5	10.1	356	19	AAW10960

12	156.5	6.6	253	18	AAW20760
13	117	5.0	1686	19	AAW70991
14	110.5	4.7	800	8	AAW70420
15	110.5	4.7	822	13	AAW26021
16	109	4.6	824	21	AAW23180
17	107	4.5	157	18	AAW20538
18	103	4.4	798	21	AAW40925
19	103	4.4	798	21	AAW90225
20	103	4.4	798	21	AAW83944
21	102.5	4.3	537	22	AAW91000
22	100.5	4.3	1726	18	AAW38756
23	98.5	4.2	467	20	AAW13378
24	98.5	4.2	467	21	AAW94860
25	98.5	4.2	467	22	AAW84601
26	98.5	4.2	467	22	AAW80246
27	96	4.1	888	22	AAW70751
28	95.5	4.0	502	22	AAW90031
29	95.5	4.0	516	22	AAW78876
30	95.5	4.0	682	17	AAW04359
31	95	4.0	1227	22	AAW81501
32	94	4.0	372	16	AAW69607
33	94	4.0	522	22	AAW95616
34	93.5	4.0	467	19	AAW75057
35	93	3.9	761	20	AAW99084
36	93	3.9	1717	22	AAW20498
37	92	3.9	600	22	AAW75092
38	92	3.9	1627	16	AAW67538
39	91.5	3.9	857	13	AAW29814
40	91.5	3.9	1024	18	AAW19604
41	91	3.9	564	21	AAW10457
42	91	3.9	564	21	AAW81952
43	90	3.8	522	22	AAW95513
44	90	3.8	790	22	AAW93045
45	90	3.8	841	20	AAW24318

ALIGNMENTS

RESULT 1

AAW85268
ID AAW85268 standard; Protein; 442 AA.

AC AAW85268;

DT 29-JUN-2000 (first entry)

XX

DE BASB034 amino acid sequence #1.

XX

KW Moraxella catarrhalis infection; BASB034; diagnosis; staging;

KW vaccine; bacteriostatic; treatment; prevention; otitis media; pneumonia;

KW sinusitis; nosocomial infection; invasive disease; chronic otitis media;

KW hearing loss; antibacterial drug.

XX

OS Moraxella catarrhalis.

XX

PN W0200015802-A1.

XX

PD 23-MAR-2000.

XX

PF 14-SEP-1999; 99WO-EP06781.

XX

PR 14-SEP-1998; 98GB-0020002.

XX

PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

XX

PI Ruelle J;

XX

DR WPL; 2000-271440/23.

XX

DR N-PSDB; AAA10700.

XX

PT Novel BASB034 polynucleotides and polypeptides from Moraxella

PT catarrhalis used to prepare vaccines against bacterial infections

H. pylori outer me
Human class II p13
Sequence encoded b
Alkaline cellulase
Bacillus sp. KSM-S
H. pylori outer me
Human ORF689
Human PGC-1 protei
Peroxisome prolife
C. glutamicum prote
Phosphatidyl inosi
Amino acid sequenc
Human protein clon
Human membrane or
Human PRO260 prote
S cerevisiae apopt
C. glutamicum prote
C. glutamicum SRT
S. pneumoniae peni
S. epidermidis ope
Gonococcal porin-5
Human protein sequ
Human secreted pro
Non-B, non-C, non-
Ostrinia nubilalis
Human colon cancer
Cytochrome p1. My
S receptor kinase
Mycoplasma genital
T. matsutake pyran
Trichoderma derivate
Human protein sequ
Human protein sequ
Mouse dephosphoryl

XX Claim 3; Fig 2; 106pp; English.
 CC This sequence represents a Moraxella catarrhalis BASB034 polypeptide from
 CC strain Mc2931 (ATCC 43617). The invention relates to BASB034 polypeptides
 CC from M. catarrhalis strains Mc2931, Mc2908, Mc2913 and Mc2969. The
 CC BASB034 polynucleotides and polypeptides may be employed as research
 CC reagents and material for the discovery of treatments and diagnostics for
 CC diseases, particularly human diseases. They are particularly used to
 CC diagnose and treat M. catarrhalis infections. They can be used for
 CC diagnosis of disease, staging of disease, or determining response of an
 CC infectious organism to drugs. The polynucleotides may be used as a source
 CC for hybridization probes, and for screening of genetic mutations,
 CC serotype, organism or strain identification, identification of mutations
 CC in BASB034 sequences, and as components of arrays which are useful for
 CC diagnostic and prognostic purposes. The polypeptides can be used to
 CC produce antibodies. The polypeptides can also be used in vaccine
 CC formulations, and to identify agonists and antagonists. The polypeptides,
 CC antibodies, agonists and antagonists (which are bacteriostatic) are used
 CC for the treatment and prevention of diseases such as otitis media in
 CC infants and children, pneumonia in elderly, sinusitis, nosocomial
 CC infections and invasive diseases, and chronic otitis media with hearing
 CC loss. The polypeptides, agonists and antagonists are also used for
 CC screening of antibacterial drugs. The BASB034 products of the invention
 CC can be used screen for new antibacterial compounds that may target
 CC resistant bacteria.
 CC
 XX
 XX Sequence 442 AA;

Query Match 100.0%; Score 2360; DB 21; Length 442;
 Best Local Similarity 100.0%; Pred. No. 4.5e-219;
 Matches 442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKVSLSTLTLSILSCFALIAIQAAVNPVAFYDEVRESENDLGDNDLPIDVOSATOSA 60
 Db 1 mkvslstltlsilscfaliaiqagavnpvafdevrsendlgdndlpdvgsatgsa 60
 QY 61 STDPANPLDEHEPELYTALENKMTLNCSALNODIMRLACYDYLVEETPAVYIKTKRSI 120
 Db 61 stdtanpldehepelytalenkmtlncsalnqdimrlacydylvgetpavlytkrsl 120
 QY 121 RLDETITWOTIKGPOVYIOETTDPIFLMGNEKGMILTKDAKOLEYAAKQFPPLSLSPFLD 180
 Db 121 rldetiwtikgpkqvlygettdpiflmngnekmltkkdaqlayaaqfplslstfld 180
 QY 181 RNNTPPLWSSRPHPNPYVLPPIFMHGKPNRSPNTPSHKAKQFTPNFERABELKFOYSVKYKA 240
 Db 181 rnnptplwssrphnpnyvlpifmhgkpnrsnptshkakqftpnferabekfyoysvkka 240
 QY 241 AEDLMGTSOGLMFGYTOOSHQIFRNGKNSRPRVNDYOPETFLQOPVYSDLPMDGKVMYI 300
 Db 241 aedlmgtsdglmfgytqshwqifrngknsrprvndyopetflqopvysdipmdgkvmyi 300
 QY 301 GMGAVHSHNGESAKISRSRNPAYILMAGMEKMLTFVMPRIKGRIFEGSGSQPDDNPDLID 360
 Db 301 gmgavhshngesaklsrsrnpayilmagmekmltfvmprikgriefegsgsqpddnplid 360
 QY 361 YYGVDVAFELVQLENKNSISGTVRYNPRSGALQLDVYVPLGKGISGYPOLFQYGGOSL 420
 Db 361 yygvdvafelvglenknsisgtvrynprrsgalqldvyvplgkgisgyfolfqyggosl 420
 QY 421 IDYHHEATSPFVGLMNDMGL 442
 Db 421 idyhheatspfvglmndmgl 442

RESULT 2
 AAY85271
 ID AAY85271 standard; Protein; 442 AA.
 XX
 AC AAY85271;
 XX

DT 29-JUN-2000 (first entry)
 XX
 DE BASB034 amino acid sequence #4.
 XX
 KW Moraxella catarrhalis infection; BASB034; diagnosis; staging;
 KW vaccine; bacteriostatic; treatment; prevention; otitis media; pneumonia;
 KW sinusitis; nosocomial infection; invasive disease; chronic otitis media;
 KW hearing loss; antibacterial drug.
 XX
 OS Moraxella catarrhalis.
 PN WO200015802-A1.
 PD 23-MAR-2000.
 XX
 PF 14-SEP-1999; 99WO-EP06781.
 XX
 PR 14-SEP-1998; 98GB-0020002.
 XX
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 PI Ruelle J;
 DR WPI: 2000-271440/23.
 DR N-PSDB: AAA10703.
 XX
 PT Novel BASB034 polynucleotides and polypeptides from Moraxella
 PT catarrhalis used to prepare vaccines against bacterial infections
 XX
 PS Claim 3; Page 69; 106pp; English.

CC This sequence represents a Moraxella catarrhalis BASB034 polypeptide from
 CC strain Mc2969. The invention relates to BASB034 polypeptides from
 CC M. catarrhalis strains Mc2931, Mc2908, Mc2913 and Mc2969. The BASB034
 CC polynucleotides and polypeptides may be employed as research reagents and
 CC material for the discovery of treatments and diagnostics for diseases,
 CC particularly human diseases. They are particularly used to diagnose and
 CC treat M. catarrhalis infections. They can be used for diagnosis of
 CC disease, staging of disease, or determining response of an infectious
 CC organism to drugs. The polynucleotides may be used as a source for
 CC hybridization probes, and for screening of genetic mutations, serotype,
 CC organism or strain identification, identification of mutations in BASB034
 CC sequences, and as components of arrays which are useful for diagnostic
 CC and prognostic purposes. The polypeptides can be used to produce
 CC antibodies. The polypeptides can also be used in vaccine formulations,
 CC and to identify agonists and antagonists. The polypeptides, antibodies,
 CC agonists and antagonists (which are bacteriostatic) are used for the
 CC treatment and prevention of diseases such as otitis media in infants and
 CC children, pneumonia in elderly, sinusitis, nosocomial infections and
 CC invasive diseases, and chronic otitis media with hearing loss. The
 CC polypeptides, agonists and antagonists are also used for screening of
 CC antibacterial drugs. The BASB034 products of the invention can be used
 CC screen for new antibacterial compounds that may target resistant
 CC bacteria.
 CC
 XX
 XX Sequence 442 AA;

Query Match 99.7%; Score 2354; DB 21; Length 442;
 Best Local Similarity 99.5%; Pred. No. 1.7e-218;
 Matches 440; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKVSLSTLTLSILSCFALIAIQAAVNPVAFYDEVRESENDLGDNDLPIDVOSATOSA 60
 Db 1 mkvslstltlsilscfaliaiqagavnpvafdevrsendlgdndlpdvgsatgsa 60
 QY 61 STDPANPLDEHEPELYTALENKMTLNCSALNODIMRLACYDYLVEETPAVYIKTKRSI 120
 Db 61 stdtanpldehepelytalenkmtlncsalnqdimrlacydylvgetpavlytkrsl 120
 QY 121 RLDETITWOTIKGPOVYIOETTDPIFLMGNEKGMILTKDAKOLEYAAKQFPPLSLSPFLD 180
 Db 121 rldetiwtikgpkqvlygettdpiflmngnekmltkkdaqlayaaqfplslstfld 180

PR 14-SEP-1998; 98GB-0020002.
 XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 XX Ruelle J;
 PI WPI: 2000-271440/23.
 XX N-PSDB: AAA10701.
 DR
 XX
 PT Novel BASB034 polynucleotides and polypeptides from Moraxella
 PT catarrhalis used to prepare vaccines against bacterial infections -
 XX
 PS Claim 3; Page 67; 106pp: English.
 XX
 CC This sequence represents a Moraxella catarrhalis BASB034 polypeptide from
 CC strain Mc2908. The invention relates to BASB034 polypeptides from
 CC M. catarrhalis strains Mc2931, Mc2908, Mc2913 and Mc2969. The BASB034
 CC polynucleotides and polypeptides may be employed as research reagents and
 CC material for the discovery of treatments and diagnostics for diseases,
 CC particularly human diseases. They are particularly used to diagnose and
 CC treat M. catarrhalis infections. They can be used for diagnosis of
 CC disease, staging of disease, or determining response of an infectious
 CC organism to drugs. The polynucleotides may be used as a source for
 CC hybridization probes, and for screening of genetic mutations, serotype,
 CC organism or strain identification, identification of mutations in BASB034
 CC sequences, and as components of arrays which are useful for diagnostic
 CC and prognostic purposes. The polypeptides can be used to produce
 CC antibodies. The polypeptides can also be used in vaccine formulations,
 CC and to identify agonists and antagonists. The polypeptides, antibodies,
 CC agonists and antagonists (which are bacteriostatic) are used for the
 CC treatment and prevention of diseases such as otitis media in infants and
 CC children, pneumonia in elderly, sinusitis, nosocomial infections and
 CC invasive diseases, and chronic otitis media with hearing loss. The
 CC polypeptides, agonists and antagonists are also used for screening of
 CC antibacterial drugs. The BASB034 products of the invention can be used
 CC screen for new antibacterial compounds that may target resistant
 CC bacteria.
 CC
 XX
 XX Sequence 442 AA;
 SQ

Query Match 98.7%; Score 2330; DB 21; Length 442;
 Best Local Similarity 98.6%; Pred. No. 3.5e-216;

Matches 436; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MKVSI STLTLISCFALIAIOAAVNPVAFVDEVSENDLGDNDLPIDVOSATOSA 60
 |||||
 Db 1 mkvslstltlslpofallaigqagavnpvrafvdevskndlgqdnelllygvaqtgsa 60
 QY 61 STDANPLDEHEPELYTALENKTM LINC S ALNODIMRLACYDTLVHGEPVAVIKTKRSI 120
 |||||
 Db 61 stdanpldehepe lytal enk tml incsalnqdlmr lac ydclvhgetpaviktkrsi 120
 QY 121 RLDETIWITIGKPOVYIOETTPDFFLMGNEKMTTKDAQOLEVAAQFIPLSIFPLD 180
 |||||
 Db 121 rlde tiw itigk pvyio ettpdf flmg ne km ttk da qole vaa qf ip ls if pl s if pl 180
 QY 181 RNNPFLMGSSRPNNPVYPIFMHGKPNRSPMPSEAKQFPNFRRAELKEOVSVYKKA 240
 |||||
 Db 181 rnnp flmg s s r p n n p v y p i f m h g k p n r s p m p s e a k q f p n f r r a e l k e o v s v y k k a 240
 QY 241 AEDLMGTDSDLMFGYTOOSHMOIFNGKNSRPERVHYDOPETFLQPVYSDLPWVGKXVMI 300
 |||||
 Db 241 aed lmg t d s d l m f g y t o o s h m o i f n g k n s r p e r v h y d o p e t f l q p v y s d l p w v g k x v m i 300
 QY 301 GCGAVHNSNGSAKISRSNRAVYLMAGMEKNLTVMPRIWGRIRKESGSGDDNDPDLTD 360
 |||||
 Db 301 gmgavhnsngs a k i s r s n r a v y l m a g m e k n l t v m p r i w g r i r k e s g s g d d n d p d l t d 360
 QY 361 YVGGVDFVFLVQLENKNSIGTVRYNPRSGGALQDLYVVLGIGISYRPIQGYGQSL 420
 |||||
 Db 361 yv g g v d f v f l v q l e n k n s i g t v r y n p r s g g a l q d l y v v l g i g i s y r p i q g y g q s l 420

QY 421 IDYNHEATSFVGIMLNDWML 442
 |||||
 Db 421 idynheatsfgv jmlndw m j 442

RESULT 5

AAV75156
 ID AAV75156 standard; Protein; 370 AA.

AAV75156;

21-MAR-2000 (first entry)

Neisseria gonorrhoeae ORF 582 protein sequence SEQ ID NO:1786.

Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;

antibacterial; gene therapy.

Neisseria gonorrhoeae.

MO9957280-A2.

11-NOV-1999.

30-APR-1999; 99WO-US09346.

01-MAY-1998; 98US-0083758.

31-JUL-1998; 98US-0094869.

02-SEP-1998; 98US-0098994.

02-SEP-1998; 98US-0099062.

09-OCT-1998; 98US-0103749.

09-OCT-1998; 98US-0103794.

09-OCT-1998; 98US-0103796.

25-FEB-1999; 99US-0121528.

(CHIR) CHIRON CORP.

(GENO-) INST GENOMIC RES.

Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M,

Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M,

Tettelin H, Venter JC;

WPI: 2000-062150/05.

N-PSDB: AA53918.

Novel Neisserial polypeptides predicted to be useful antigens for

vaccines and diagnostics -

Claim 2; Page 903; 1453pp: English.

AA53015 to AA54536, AA54577 to AA54615, and AAV74253 to AAV75941

represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides

and polypeptides. AA54537 to AA54576 and AA54616 to AA54673 represent

PCR primers used in the exemplification of the present invention. The

polynucleotides, antibodies and compositions of the present invention can

be used as vaccines, as diagnostic reagents, and as

immunogenic compositions. The polypeptides can be used in the

manufacture of medicaments for treating or preventing infection due to

Neisserial bacteria (e.g. meningitis and septicaemia), to detect the

presence of Neisseria bacteria, or to raise antibodies. They may also

be used to screen for agonists or antagonists, which may themselves

have use as antibacterial agents. The polynucleotides of the invention

may also be used in gene therapy protocols.

Sequence 370 AA;

Query Match 34.8%; Score 821; DB 21; Length 370;

Best Local Similarity 44.4%; Pred. No. 1.2e-70;

Matches 162; Conservative 69; Mismatches 110; Indels 24; Gaps 8;

QY 87 INC S ALNODIMRLACYDTLVHGEPVAVI-----KTKRSIRLDETITWQTI-KGKPOVYIOE 140


```

Db      21 lqcaaltvntlracydrifaagqlpsasagqegqeskaavlnctetvrsldkgeavlvvek 80
Qy      141 TTDPPIFLMGNEKGMKTKKAKOLEYAAKQFTPLSLSPFDLDRNN-TPLMSSRPNNPVLP 199
Db      81 ggda1-----padsagetadlytpislmydldkndlrgllgyrehnmpyimp 127
Qy      200 IFMHGKPNRSPNTPSH-EAKQFTPNFRAPELKFQYSVKVKAEDLMGTSDDLMEFGYTQO 258
Db      128 lymnspnyapgsprtgltvqekfqqktraetklysfksklaenlftktradiwfygtqr 187
Qy      259 SHWQIFN-GKNSRPFVHDYQPEIFLTQPVYSDLPMDGKVRMIGMGAHNSGESAKLSR 317
Db      188 sdwqlyngqrksapfrntdykpeliftqpvkadlpfggrlrmagafvbnsgsqsrpear 247
Qy      318 SMNRAYLMAGMEKKNLTVMPRIMGRIFKESGSGQPDNDPILDYGYGVRFLYOLENKS 377
Db      248 swrilyamagmewgkltvprvvratfdq-sgdk-ndnpdiadymgydvlkqyrindrq 305
Qy      378 NISGTVRYNPRSGKALOLDYVYPLGKISGYFOIFQYGSGLIDYNHEATSPGVGLMLN 437
Db      306 nysvvlrnpktygataeaaytflpkyklkgyvrgfngygeslldynhknqnglqldmfn 365
Qy      438 DMWGL 442
Db      366 dwdgl 370

```

```

RESULT 6
AA75157
AA75157 standard; Protein; 370 AA.
AC  AA75157;
DE  21-MAR-2000 (first entry)
XX  Neisseria meningitidis ORF 582 protein sequence SEQ ID NO:1788.
XX  Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW  antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX  antibacterial; gene therapy.
XX  Neisseria meningitidis.
OS  Neisseria meningitidis.
PN  WO957280-A2.
XX  11-NOV-1999.
XX  30-APR-1999; 99WO-US09346.
XX  01-MAY-1998; 98US-0083758.
PR  31-JUL-1998; 98US-0094869.
PR  02-SEP-1998; 98US-0098994.
PR  02-SEP-1998; 98US-0099062.
PR  09-OCT-1998; 98US-0103749.
PR  09-OCT-1998; 98US-0103794.
PR  09-OCT-1998; 98US-0103796.
PR  25-FEB-1999; 99US-0121528.
XX  (CHIR) CHIRON CORP.
PA  (GENO-) INST GENOMIC RES.
XX  Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI  Petersen J, Pizsa M, Rappoli R, Ratti G, Scalato E, Scarselli M;
PI  Tettelin H, Venter JC;
XX  WPI; 2000-062150/05.
DR  N-PSDB; AA53919.
XX  Novel Neisserial polypeptides predicted to be useful antigens for
PT  vaccines and diagnostics
XX  Claim 2; Page 904; 1453pp; English.

```

```

XX  AA53015 to AA254536, AA254577 to AA254615, and AA774253 to AA75941
CC  represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC  and polypeptides. AA254537 to AA254576 and AA254616 to AA254673 represent
CC  PCR primers used in the exemplification of the present invention. The
CC  polypeptides, the polynucleotides, antibodies and compositions of
CC  the invention can be used as vaccines, as diagnostic reagents, and as
CC  immunogenic compositions. The polypeptides can be used in the
CC  manufacture of medicaments for treating or preventing infection due to
CC  Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC  presence of Neisseria bacteria, or to raise antibodies. They may also
CC  be used to screen for agonists or antagonists, which may themselves
CC  have use as antibacterial agents. The polynucleotides of the invention
CC  may also be used in gene therapy protocols.
SQ  Sequence 370 AA:

```

```

Query Match      34.5%; Score 814; DB 21; Length 370;
Best Local Similarity 44.4%; Pred. No. 5 9e-70;
Matches 162; Conservative 68; Mismatches 111; Indels 24; Gaps 8;

```

```

Qy      87 INCSALNODIMRLACYDTLVHGETPAVI-----KTKRSIRLDEFTIMQTI-KGRPOVITYOE 140
Db      21 lqcaaltvntlracydrifaagqlpsasagqegqeskaavlnctetvrsldkgeavlvvek 80
Qy      141 TTDPPIFLMGNEKGMKTKKAKOLEYAAKQFTPLSLSPFDLDRNN-TPLMSSRPNNPVLP 199
Db      81 ggda1-----padsagetadlytpislmydldkndlrgllgyrehnmpyimp 127
Qy      200 IFMHGKPNRSPNTPSH-EAKQFTPNFRAPELKFQYSVKVKAEDLMGTSDDLMEFGYTQO 258
Db      128 lymnspnyapgsprtgltvqekfqqktraetklysfksklaenlftktradiwfygtqr 187
Qy      259 SHWQIFN-GKNSRPFVHDYQPEIFLTQPVYSDLPMDGKVRMIGMGAHNSGESAKLSR 317
Db      188 sdwqlyngqrksapfrntdykpeliftqpvkadlpfggrlrmagafvbnsgsqsrpear 247
Qy      318 SMNRAYLMAGMEKKNLTVMPRIMGRIFKESGSGQPDNDPILDYGYGVRFLYOLENKS 377
Db      248 swrilyamagmewgkltvprvvratfdq-sgdk-ndnpdiadymgydvlkqyrindrq 305
Qy      378 NISGTVRYNPRSGKALOLDYVYPLGKISGYFOIFQYGSGLIDYNHEATSPGVGLMLN 437
Db      306 nysvvlrnpktygataeaaytflpkyklkgyvrgfngygeslldynhknqnglqldmfn 365
Qy      438 DMWGL 442
Db      366 dldgl 370

```

```

RESULT 7
AA75158
AA75158 standard; Protein; 370 AA.
AC  AA75158;
DE  21-MAR-2000 (first entry)
XX  Neisseria meningitidis ORF 582 protein sequence SEQ ID NO:1790.
XX  Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW  antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX  antibacterial; gene therapy.
XX  Neisseria meningitidis.
OS  Neisseria meningitidis.
PN  WO957280-A2.
XX  11-NOV-1999.
XX  30-APR-1999; 99WO-US09346.

```

```

PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
PI Petersen J, Pizzer M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI; 2000-062150/05.
DR N-PSDB; AA53920.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
XX
XX Claim 2; Page 905; 1453pp; English.
XX
XX AA53015 to AA54536, AA54577 to AA54615, and AA74253 to AA75941
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX and polypeptides. AA534537 to AA54576 and AA54616 to AA55473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisseria meningitidis (e.g. meningitis and septicemia), to detect the
XX presence of Neisseria meningitidis, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX
XX Sequence 370 AA:
SO
Query Match 34.5%; Score 814; DB 21; Length 370;
Best Local Similarity 44.4%; Pred. No. 5,9e-70;
Matches 162; Conservative 68; Mismatches 111; Indels 24; Gaps 8;
QY 87 INCSALNODIMRLACYPDLVHGEPYAVI---KTKRSIRLDEITMOTI-KGKPOVIYOE 140
DB 21 lqcaaltndvtrclaeaydrilfaaqlpsasgqegskavlnltetrssldkgeavilvek 80
QY 141 TTDPFTLGNENEGMLTKKDAKOLEYAAKOTFPLSLFPLDRNN-TPLMSSRPHNMYVLP 199
DB 81 ggdaL-----padsagetadlyrplslmyaldkndlrgllgvrhpnmylmp 127
QY 200 IFMHGKPNRSPNTPSH-BAKOTFNEFRAPELKPOVSVKVAABEDMTSDSLMGTYOQ 258
DB 128 lwmnspnyapgsprgtctvgekfgqkkaetklyvstkskaedlftktradlwfygqr 187
QY 259 SHWOIFN-GKNSRPRVHDYQPEIFLTOPYVSDLPWDGKVRMIGAVHNSGESAKLSR 317
DB 188 sdwqlynggrksapfrndykpelftlcpvkadlprfggrlmlgagfthgngsgqrpsr 247
QY 318 SMNRATYLMAGMEKMLTYMPRIWGRIFKESGSDPDNDPDLIDYGYGDVDFRFLYLEKMS 377
DB 248 swrtlyamagmewgkltvlpvrvwrafq-sgdk-ndnprdladyngygdvdklqyrlndrq 305
QY 378 NISGTVRYNPNRSGKALDLVDYVPLGKIGISGYFOIFGOGSLIDYNEATSFVGMLN 437
DB 306 nvsyrlrnpktygaiaaaylfpklgklkvrvvgfhyggesldylnkqngidgldmfn 365
QY 438 DWGML 442
DB 366 dlldgl 370

```

```

RESULT 8
AA70629
ID AA70629 standard; Protein; 374 AA.
XX
XX AA70629;
AC
XX 18-JUL-2000 (first entry)
XX
XX Neisseria meningitidis serogroup B strain H44/76 BASB033 protein.
DE
XX BASB033; diagnosis, prophylaxis; treatment; antibacterial; vaccine;
XX Neisseria meningitidis infection.
KW
XX Neisseria meningitidis.
OS
XX WO200015801-A1.
XX
XX 23-MAR-2000.
XX
XX 09-SEP-1999; 99WO-EP06718.
XX
XX 14-SEP-1998; 98GB-0020003.
XX
XX (SMIT ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Ruelle J;
XX
XX WPI; 2000-271439/23.
XX
XX N-PSDB; AA52134.
XX
XX Isolated BASB033 polypeptides and polynucleotides of Neisseria
XX meningitidis, useful for diagnosis, prophylaxis and treatment of N.
XX meningitidis infection
XX
XX Claim 4; Page 59; 93pp; English.
XX
XX The present sequence is a BASB033 protein from
XX Neisseria meningitidis serogroup B strain H44/76. The protein
XX shows homology to the Klebsiella pneumoniae outer membrane
XX phospholipase A. The present sequence is useful for diagnosis,
XX prophylaxis and treatment of N. meningitidis infection. It may also be
XX used for the discovery and development of antibacterial compounds and
XX in vaccine compositions.
XX
XX Sequence 374 AA:
SO
Query Match 34.5%; Score 814; DB 21; Length 374;
Best Local Similarity 44.4%; Pred. No. 6e-70;
Matches 162; Conservative 68; Mismatches 111; Indels 24; Gaps 8;
QY 87 INCSALNODIMRLACYPDLVHGEPYAVI---KTKRSIRLDEITMOTI-KGKPOVIYOE 140
DB 25 lqcaaltndvtrclaeaydrilfaaqlpsasgqegskavlnltetrssldkgeavilvek 84
QY 141 TTDPFTLGNENEGMLTKKDAKOLEYAAKOTFPLSLFPLDRNN-TPLMSSRPHNMYVLP 199
DB 85 ggdaL-----padsagetadlyrplslmyaldkndlrgllgvrhpnmylmp 131
QY 200 IFMHGKPNRSPNTPSH-BAKOTFNEFRAPELKPOVSVKVAABEDMTSDSLMGTYOQ 258
DB 132 lwmnspnyapgsprgtctvgekfgqkkaetklyvstkskaedlftktradlwfygqr 191
QY 259 SHWOIFN-GKNSRPRVHDYQPEIFLTOPYVSDLPWDGKVRMIGAVHNSGESAKLSR 317
DB 192 sdwqlynggrksapfrndykpelftlcpvkadlprfggrlmlgagfthgngsgqrpsr 251
QY 318 SMNRATYLMAGMEKMLTYMPRIWGRIFKESGSDPDNDPDLIDYGYGDVDFRFLYLEKMS 377
DB 252 swrtlyamagmewgkltvlpvrvwrafq-sgdk-ndnprdladyngygdvdklqyrlndrq 309
QY 378 NISGTVRYNPNRSGKALDLVDYVPLGKIGISGYFOIFGOGSLIDYNEATSFVGMLN 437

```

DB 310 nysvslrnpkctgygaieaayctfplkglkgyvrgfhygseildqnhkngqigldmfn 369
 QY 438 DWMGL 442
 DB 370 dldgl 374

RESULT 9

AAV70628
 ID AAV70628 standard; Protein; 375 AA.

AAV70628;

18-JUL-2000 (first entry)

Neisseria meningitidis serogroup B strain ATCC13090 BASB033 protein.

BASB033; diagnosis, prophylaxis; treatment; antibacterial; vaccine;

Neisseria meningitidis infection.

Neisseria meningitidis.

WO200015801-A1.

23-MAR-2000.

09-SEP-1999; 99WO-EP06718.

14-SEP-1998; 98GB-0020003.

(SMK) SMITHKLINE BEECHAM BIOLOGICALS.

Ruelle J;

WPI; 2000-271439/23.

DR N-PSDB; AA52133.

Isolated BASB033 polypeptides and polynucleotides of Neisseria meningitidis, useful for diagnosis, prophylaxis and treatment of N. meningitidis infection.

Claim 4; Page 58; 93pp; English.

The present sequence is a BASB033 protein from Neisseria meningitidis serogroup B strain ATCC13090. The protein shows homology to the Klebsiella pneumoniae outer membrane phospholipase A. The present sequence is useful for diagnosis, prophylaxis and treatment of N. meningitidis infection. It may also be used for the discovery and development of antibacterial compounds and in vaccine compositions.

Sequence 375 AA;

Query Match 34.4%; Score 811; DB 21; Length 375;

Best Local Similarity 44.1%; Pred. No. 1,2e-69;

Matches 161; Conservative 69; Mismatches 111; Indels 24; Gaps 8;

QY 87 INCSALNODIMRLACYDTLVNGETPAVI-----KTKRSIRLDETIWOTI-KGRPOYIOE 140

DB 26 lqcaeltlnvtlvcydrilfaeqilpsagsgqsgskavlnltctvssldkgaavlvvek 85

QY 141 TTDPIFLMGNEKGMILTAKKOLEYAAKOPTPLISFSDIDRNN-TPLWSSRPHNMYVLP 199

DB 86 ggdal-----pdsagetadlytprlslnydlidkndrlgllgyrehnpylmp 132

QY 200 IEHNGKPNPSNPSH-EAKQFTPNFEFARPELKEQVSVKYKAEDLMGTDSDLMFGYTOQ 258

DB 133 lwnspnpyapsprlgtctvqekfgqktaeklygsfksklaedlftktradiwfgytr 192

QY 259 SHMOJFN-GKNSRPFVHDYQPEIFLTQPVYSDLPMDGKRMIGMCAVHNSGESAKLRF 317

DB 193 sdqylnqgrksafrntdykpeliltqpvkadielfggrilrmigagfvbqngsqsrpeet 252
 QY 318 SWNRATLMAEMENKNTWPRIRGIRFKEGSGSQPDNDPIIDYGYGQVRFYOLENKS 377
 DB 253 swnrlyamagmewgkltlvprvvralfdq-sgdk-ndnpladymgygdqklyrindrq 310
 QY 378 NISGTVRNPNRSGKALQLDYVYPLKGISGYFOIFOGYGOSLIDYNEHETSFGVGLMLN 437
 DB 311 nysvslrnpkctgygaieaayctfplkglkgyvrgfhygseildqnhkngqigldmfn 370

QY 438 DWMGL 442

DB 371 dldgl 375

RESULT 10

AAW98871
 ID AAW98871 standard; Protein; 355 AA.

AAW98871;

31-MAR-1999 (first entry)

H. pylori GHP0 1723 protein.

GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;

peptic ulcer disease.

Helicobacter pylori.

WO9843478-A1.

08-OCT-1998.

01-APR-1998; 98WO-US06371.

29-JUL-1997; 97US-0902615.

PR 01-APR-1997; 97US-0833457.

PR 24-JUN-1997; 97US-0881227.

(HUMA-) HUMAN GENOME SCI INC.

PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

PI Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;

DR WPI; 1998-542293/46.

DR N-PSDB; AAX14590.

New isolated Helicobacter polynucleotides - used to develop products for the diagnosis, prevention and treatment of Helicobacter infections and gastrointestinal diseases

Claim 8; Page 1976-1977; 2054pp; English.

This sequence represents a Helicobacter pylori GHP0 protein of the CC invention. The polypeptides can be used for preventing or treating CC Helicobacter infections, and gastroduodenal diseases associated with CC these infections, including acute, chronic, and atrophic gastritis, and CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be CC used for the production of antibodies. The products can also be used for CC detection and diagnosis.

Sequence 355 AA;

Query Match 10.4%; Score 246.5; DB 19; Length 355;

Best Local Similarity 25.1%; Pred. No. 3.4e-15;

Matches 82; Conservative 41; Mismatches 117; Indels 87; Gaps 11;

QY 157 KDAKOLEYAAKOPTPLISFSDIDRNNTPLMSSRPHNMYVLPDIFHNGKPNPSNPSH 216

DB 69 kkylnmndylygtvflpfyhsf-----tpifqwyhpninp----- 102

PF 02-APR-1987; 87JP-0076285.
 XX 02-APR-1986; 86JP-0076285.
 XX (RIKA) RIKAGAKU KENKYUSHO.
 PA WPI: 1987-325245/46.
 XX N-PSDB: AAN70651.
 DR
 XX DNA sequence coding cellulase gene - is derived from *Bacillus* sp. no.
 PT 1139z and can hydrolyse cello-triose and cello-tetraose.
 XX
 XX Disclosure: Fig 3-1 - 3-3; 11pp; Japanese.
 PS
 CC This cellulase hydrolyses cellobiose or cellobiose and does not
 CC hydrolyse cellobiose. Molecular weight is approx. 92K dalton and
 CC optimum pH is 9.
 XX
 SQ Sequence 800 AA:

Query Match 4.7%; Score 110.5; DB 8; Length 800;
 Best Local Similarity 20.0%; Pred. No. 0.16;
 Matches 96; Conservative 53; Mismatches 163; Indels 169; Gaps 25;

QY 9 TLTSTSCFAILLAIQQAQAVPNVAFVDEVRSND---LGQDNELPIDVQASTQASTDTA 65
 DB 7 tkqlissalllvtl-llstlftalaegnltrednfkhllgndvkrpseagallqgevqg 65
 QY 66 NPL-DEHEPELY-----TTALENKTMLINCAL-----NODIMRLACY-DTLVHGEP 111
 DB 66 mlivqghekiqlrgmsthtgqlwfpellndnaykalandwesnmrlamygengyasnp 125
 QY 112 AVIKTKRSIRLDETI-----WQT-----IKGKPOVLY 138
 DB 126 elikervlkgldlaldendmyvlvdvhnapgqprdyagaedffrdlaalympnhilly 185
 QY 139 QETTPP-----IFLMGNEKMLTKKDAKOLEYAAKQPTPLSFSFDLDRN---NTPLM 187
 DB 186 elanepssnmnggaglpnneegwaavk---eyadpivemlrdsngadnllivgsnpw 240
 QY 188 SSRP-----HNPMVLPPIF--MHGKPNRS--PMTPSHEAKQFPNEFRAPELKFO 233
 DB 241 sqrpolaadnpridhthmytvtfhfysghaastesyppepnsnrgnvnmtlyta---le 296
 QY 234 VSVKKAEDLMGT-----DSDLMPGYTQO-----SHWOIFNGKNS-----RP 271
 DB 297 ngvavfate--wgtseqangdgapyfdeadvwleflnemlswanwsltn-knevsqaftr 353
 QY 272 FRVH-----DYOPEIFLTOPVYSDLPM---D 294
 DB 354 felgksnatlsldpdpdqvwppeelslsggyvratikgyvnyep---ldtklytkvldwdfnd 410
 QY 295 GKVRIAGAVHSHNGES-----AKLSRSMNRAYLMAGMEKMLNLTVPRIWG 341
 DB 411 gtkgqfgyv-----ngdspvedvvieneagalklsgldasndvsegnynwanarlsadgwg 464
 QY 342 R 342
 DB 465 k 465

RESULT 15
 ID AAR26021 standard; Protein: 822 AA.
 AAR26021;
 02-FEB-1993 (first entry)
 Alkaline cellulase K-64.
 Mass production; recombinant.

XX *Bacillus* sp. KSM-64 (FERM P-10482).
 OS JP04190793-A.
 PN
 XX 09-JUL-1992.
 PD
 XX 26-NOV-1990; 90JP-0324401.
 PF 26-NOV-1990; 90JP-0324401.
 XX
 XX 26-NOV-1990; 90JP-0324401.
 PR (KAOS) KAO CORP.
 XX WPI: 1992-280112/34.
 DR N-PSDB: AAO27180.
 XX
 PT Alkaline cellulose mass prodn. - using recombinant plasmid and
 PT microorganism contg. cellulase gene
 PS Disclosure: Fig 4; 14pp; Japanese.

Query Match 4.7%; Score 110.5; DB 13; Length 822;
 Best Local Similarity 20.0%; Pred. No. 0.17;
 Matches 96; Conservative 53; Mismatches 163; Indels 169; Gaps 25;

QY 9 TLTSTSCFAILLAIQQAQAVPNVAFVDEVRSND---LGQDNELPIDVQASTQASTDTA 65
 DB 7 tkqlissalllvtl-llstlftalaegnltrednfkhllgndvkrpseagallqgevqg 65
 QY 66 NPL-DEHEPELY-----TTALENKTMLINCAL-----NODIMRLACY-DTLVHGEP 111
 DB 66 mlivqghekiqlrgmsthtgqlwfpellndnaykalandwesnmrlamygengyasnp 125
 QY 112 AVIKTKRSIRLDETI-----WQT-----IKGKPOVLY 138
 DB 126 elikervlkgldlaldendmyvlvdvhnapgqprdyagaedffrdlaalympnhilly 185
 QY 139 QETTPP-----IFLMGNEKMLTKKDAKOLEYAAKQPTPLSFSFDLDRN---NTPLM 187
 DB 186 elanepssnmnggaglpnneegwaavk---eyadpivemlrdsngadnllivgsnpw 240
 QY 188 SSRP-----HNPMVLPPIF--MHGKPNRS--PMTPSHEAKQFPNEFRAPELKFO 233
 DB 241 sqrpolaadnpridhthmytvtfhfysghaastesyppepnsnrgnvnmtlyta---le 296
 QY 234 VSVKKAEDLMGT-----DSDLMPGYTQO-----SHWOIFNGKNS-----RP 271
 DB 297 ngvavfate--wgtseqangdgapyfdeadvwleflnemlswanwsltn-knevsqaftr 353
 QY 272 FRVH-----DYOPEIFLTOPVYSDLPM---D 294
 DB 354 felgksnatlsldpdpdqvwppeelslsggyvratikgyvnyep---ldtklytkvldwdfnd 410
 QY 295 GKVRIAGAVHSHNGES-----AKLSRSMNRAYLMAGMEKMLNLTVPRIWG 341
 DB 411 gtkgqfgyv-----ngdspvedvvieneagalklsgldasndvsegnynwanarlsadgwg 464
 QY 342 R 342
 DB 465 k 465

Search completed: November 30, 2001, 14:16:58
 Job time: 380 sec

```

: ATTORNEY/AGENT INFORMATION:
:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 39116-A
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: (212) 278-0400
:
: TELEFAX: (212) 391-0525
:
: INFORMATION FOR SEQ ID NO: 6:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 985 amino acids
:
: TYPE: amino acid
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: PCT-US96-03916-6

```

Query Match 3.7%; Score 87; DB 5; Length 985;
Best Local Similarity 19.4%; Pred. No. 6.1;
Matches 73; Conservative 47; Mismatches 130; Gaps 20

```

QY      48 ELRPDVQASQASQSTANPLADHEPELYT-TALENTKTMLNCSALNODIMTACDITLV 106
Db      563 QTPSTVKRTLLISST-----EPALFRTQASGAEATQSSAPBDIMRQSTETHR 613
QY      107 HGPMPAVIKTKRSIRLDEITWQTIKGRQVUYOE-TTDPI--FLAMENKMLTKKDAKO 162
Db      614 FTQAPSTVPKAT-----QTPSTPEVLTQSPSTPEVPFRTIGAEPEIQTPSA-- 662
QY      163 LEYAAKQFTPLSLSFDDDRNNTPLMSSRP-----HN-----PMYVLP---IFNHCK 205
Db      663 ---APEVYTRSSSTMPETAOSTPLASQNPSSGSGHTNTPRTYPVQTPPHQKLTENK 719
QY      206 PNRSPNPS--HE--AKQFPR-NEPARPELKQVSVKYKA-----AEDLMGTD 248
Db      720 TLSEPTVYSEFHEMSTAESQTPRLLDVKTVEKFKSSNDGEVATATCVSVYKSPYRVEITMKVD 779
QY      249 -----SDLMFGYQOOSHQOIFENKNS-----RPF----- 272
Db      780 LVQVMDIELSGNSPAGVFNSENKK---QKOLYLRVTDQRTSVQMLCLSGTSHSPPECLFD 836
QY      273 -----RVHNDQPELEFL---QPVYSDLPMDGKY-----RMIGGAVHNHN 309
Db      837 TSLIAREKDIAPBELYFTSSDQATCYTLRSGVVRPREMSLNNVSLPEVLTATTVVSHTA 896
QY      310 GE-----SAKLSRSM 319
Db      897 GQSTVWKSSSAPAGEAW 912

```

RESULT 13
PCT-US96-03916-66
Sequence 66: Application PC/TUS9603916
GENERAL INFORMATION:
APPLICANT: Wild, Martha A.
APPLICANT: Cochran, Mark D.
TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS
TITLE OF INVENTION: AND USGS THEREOF
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/03916
FILING DATE: 23-MAR-1995

```

1 CLASSIFICATION:
2 PRIOR APPLICATION DATA:
3 APPLICATION NUMBER: US 08/126,597
4 FILING DATE: 24-SEP-1993
5 ATTORNEY/AGENT INFORMATION:
6 NAME: White, John P.
7 REGISTRATION NUMBER: 28,678
8 REFERENCE/DOCKET NUMBER: 39116-A
9 TELECOMMUNICATION INFORMATION:
10 TELEPHONE: (212) 278-0400
11 TELEFAX: (212) 391-0525
12 INFORMATION FOR SEQ ID NO: 66:
13 SEQUENCE CHARACTERISTICS:
14 LENGTH: 985 amino acids
15 TYPE: amino acid
16 TOPOLOGY: linear
17 MOLECULE TYPE: protein
18 PCT-US96-03916-66

```

Query Match	3.7%;	Score 87;	DB 5;	Length 985;
Best Local Similarity	19.4%;	Pred. No. 6.1;		
Matches	73;	Conservative	47;	Mismatches 126;
				Indels 130;
				Gaps 20;

```

OY      48  ELPIDVOSATOSADPNANPLDEHEPELYT--TALENTKTMILNCSALMODIMLACDYTLV 106
Db      563  QTPSPVPEKTTILSSST-----EPALFTTQSGAGTEATQISSAPDPMRQSIETHF 613
OY      107  HGETPAVYKTRKSIRILDETTWOTIKGAPVOUYOB--TDPPI--FLMGNEKMLTKRDAKO 162
Db      614  FTQAPSTVPKAT-----QTPSTDEVELTQSPSTPEVPFTTIGAAPELITQTPSA-- 662
OY      163  LEYAAKOTTPILSLSFDDDRANTPLMSSRP-----HN-----PMYVLP-----IFMHCK 205
Db      663  --APEVYTRBSSTMPETAOSTPLASONPPTSSGCTGTNTEPRTPYVQTPPHOKLYTENK 719
OY      206  PNRSPNPS--HE--AKQFPR--NEFAPLEKFCVSKYKA-----AEDLMGTD 248
Db      720  TLSFPTVYSEFHEMSTAESQTPLLDVKIVEYKRSNDGEVATATCVSYVSKSPYRVELTMKKVD 779
OY      249  -----SDLMFGYTOQSHWOJFNCKNS-----RPF----- 272
Db      780  LVDYVMEIDSGNSPAGVFNSENKMK--QKOLYRYTDRTERTSVQMLCLSCSTSHSPEPCLFD 836
OY      273  -----RHADYQPELFLT--QPVYSDLPMDCKV-----RMIGGAYVHHSN 309
Db      837  TSLIAREKDILPELYFTSDPOTACYTILPSGVVPREEMSLNNVSLPELYLTATTVVSHTA 896
OY      310  GE-----SAKLSRWS 319
Db      897  GOSTVWKKSSARAGAEAW 912

```

RESULT 14
 US-08-961-083-2
 : Sequence 2, Application US/08961083
 : Patent No. 6158469
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Choi et. al.
 : TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 :
 : NUMBER OF SEQUENCES: 452
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Human Genome Sciences, Inc.
 : STREET: 9410 Key West Avenue
 : CITY: Rockville
 : STATE: Maryland
 : COUNTRY: USA
 : Zip: 20850
 :
 : COMPUTER READABLE FORM:
 :
 : MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 : COMPUTER: HP Vectra 486/33
 : OPERATING SYSTEM: MSDOS version 6.2
 : SOFTWARE: ASCII Text

Best Local Similarity 21.18; Pred. No. 2.7;
Matches 75; Conservative 51; Mismatches 110; Indels 119; Gaps 23

```
OY      105 LVHGEPAVITKRSIRDEIIMQTINGKPOVUIQEETDPELMG-----NEK-----GWLTk 157
Db      24 LHPALSIYIMT-----LSSTESLISSNKTLV---SPGSIFEEGFERTSRMYLGMMYK 75
OY      158 K-DAKOLEYAAKOFTEPLSLSPD-----LDRNNTPPLMSSRPHPNPMIVLPILFMHGK 205d
Db      76 KVSDRITYWVANRDNPLESNALGTIKSIGNNLVLLDDHSHKNKPWWMTN-----LTRGN 125d
OY      206 PNRSD-----NTPSHEAKOFT-----PNEFRAPELKFQSVKYKAEDLW 245
Db      126 -ERGVVAELLANGNFVAROSSNNDASEYLQMOSFDYTPTDLRLPRMKIGYNLKT----- 177d
OY      246 GTDSDDLMEGYTOOSHWOIFENGSKNSRPVRHYDQ-----PEILFLO--PVUSDLPMDGK 296
Db      178 GLNRFL-----TSWRSDDPSSGNGF---SYKLETOGLPEFLYSRNPNRHSGRPWGNG- 226d
OY      297 VRATGMCAVH-----HSNESAKLSISNMURAY-----LMAGEMKENLYMP--RI 339d
Db      227 IRFGSIPEDOKLSVMVYNFIENNEEVAVYTFEMTNNSFSRLTILISEGYFORLWTWPSPSIRI 286d
OY      340 WGRJFKKESSGSOPOD---NPDLIDYGYGGVRRLYOLEKMSNISGVLR-YNPRS 389
Db      287 WNRFWSSVPVDROCTPIYMGP---YAYICV-----NTSPVCNCIOGPNPN 329
```

RESULT 11
US-08-633-879C-2

1 GENERAL INFORMATION:
 2 APPLICANT: Kivirikko, Karl I.
 3 APPLICANT: Pihlajaniemi, Taina
 4 APPLICANT: Helakoski, Tarja I.
 5 APPLICANT: Annunen, Pia P.
 6 APPLICANT: Nissi, Riitta K.
 7 APPLICANT: NO. 592892elainen, Minna K.
 8 TITLE OF INVENTION: 2 SUBUNIT OF PROLYL-4-HYDROXYLASE
 9 TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING SUCH SUBUNIT AND
 10 METHODS FOR PRODUCING THE SAME
 11 NUMBER OF SEQUENCES: 20

```

;      TYPE: amino acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      FRAGMENT TYPE: internal
US-08-633-879C-2

```

Query Match	3.7%;	Score 87;	DB 2;	Length 537;
Best Local Similarity	21.0%;	Pred. No. 2.3;		
Matches 98;	Conservative 53;	Mismatches 146;	Indels 170;	Gaps 25

QY	1	MKVSJLSTLTLISCFALALAOQAQAVNPAPAFDEVRSENDJGQDNELPIDVQSA	56
Db	1	MKLOVLVYLVLMSMFYGLSMVQAEFTTSGHMTDLYAEKDLVQSLKEVTLVBEALAK	59
QY	57	-----TQASASTD-----TANPLDEHE-----PELYTALENKT-----MLINCS	90
Db	60	IKSMASKMEALTSTSSADDEGYLAHPNVAKYKRLMTDWPALGDVLIDQASAGFVNLS	119
QY	91	ALNQ-----DIPMLA-CY-----DTLVHGETPRAVIKTKRSIRDERI-----	126
Db	120	VQRFPTDESEGAARALMKLQDTYKILDPDTISRGLPFGT-KYQAMLVDCCFGJGRSA	178
QY	127	-----MOTIKRQPVLYQFETTDPIFLMGNKMGMLTKR-DAKOLEYAAKOFPLTSLFSL	179
Db	179	YNEGDYHTTVLTMQOVLKQ-----LDAGEAAYTTSVLVDLYSTAVPQGLDHRARVEL	231
QY	180	DRNNTPLMSSRPNNPMTVLLPIEMHGKPNRSPNTPRSEAK-----QFTNERRAPRLAKT	232
Db	232	TJRLRLSTL-----PESHERRAGGNLRYFERLLLEEGKSLSN	266
QY	233	QVSVKVAKAEDJLWGTDSJLWFCYTOQSHWQJFNKNSRPRVHNDYOEIFLTPOVYSDLP	292
Db	267	QTDAGLATQEWLY-----ERP-----TDYLDE-----ROYESL-	295
QY	293	WDGKRYMIGMAVHNHNGESAKLS-RSMNRYALTM--AGMEMKILTVMPRIAGRIFKEGSG	349
Db	296	-----CQGEQVKYLPRRQKKLFCYUHNHGNVRPOLLNAP-----FKR-----	351
QY	350	SQPDNDPDIIDLYGCGVDFRLYQLEKNSNLS-----GYRATNPRSG	390
Db	332	EDWDSPIVIRYVY--DVMSDEDETERIKELAKPRKLARATVR-DPKTG	374

STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2811
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/633,879C
FILING DATE: 10-APR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Abrams, Samuel B
REGISTRATION NUMBER: 30,605
REFERENCE/DOCKET NUMBER: 8389-0041-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 537 amino acids

RESULT 12
PCT-US96-03916-6
Sequence 6, Application PC/TUS9603916
GENERAL INFORMATION:
APPLICANT: Wild, Martha A.
APPLICANT: Cochran, Mark D.
TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOOS
SOFTWARE: Patent In Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/03916
FILING DATE: 23-MAR-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,597
FILING DATE: 24-SEP-1993

RESULT 9
US-08-478-882-2
; Sequence 2, Application US/08478882
; Patent No. 5885575

```

Query Match          3.8%; Score 89.5; DB 2; Length 503;
Best Local Similarity 18.4%; Pred No.1.1;
Matches 57; Conservative 32; Mismatches 96; Indels 125; Gaps 14;

OY 40 ENLDGQDELPIDVOSATQASASTDTPANPLDEHEPELYTALLENKTMILNCALNODIMRL 99
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 169 QGDIDASNITDEADVSS-----GSTIEKSTPGEGILHTDLPT--SQPGDRD----- 212
OY 100 ACYDPLVHGEPYAVIKTKRST-----RLDEITWOTIKGPGVITYQETDPLFLMGNENG 153
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 213 --DAFFIGSTLATIATTPWVSANHTKQNERQWMPHISNPVELLQTYT----- 258
OY 154 MLTKKDAKOLEAAQKFTPLSLSPDLDRNNTPL---WSSRPHNPMYVLPIMHGKPNRS 209
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 -----RMT-----DIDRSTSAHGEMTQGEQPPF----- 283
OY 210 PNTPEHKAQKFTPNNEPRAPELKFOVSVKYKAEDLW-----GTDSDLMFGITQOS 259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 284 -NNHEQOEDEETPR-----ASTTWADPNSITTEEAATQKEKWF-----EN 322
OY 260 HMOINGKN-----SRPRVHDYOPEIFLTQPVYSDLPW---DGKVRMIGMG 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 EMO---GKKPPPSSEDSHTEGTITASAHNNHPSQRRTTOSQDSQVWTFDFDPSHPMGOG 379
OY 304 AVHHSNGESA 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 380 HOTESKGHSS 389

RESULT 10
US-07-717-331F-2
; Sequence 2, Application US/07717331F
; Patent No. 5484305
; GENERAL INFORMATION:
; APPLICANT: June Nasrallah; Michael Nasrallah; and Joshua
; APPLICANT: Stein
; TITLE OF INVENTION: A Receptor Protein Kinase Gene
; TITLE OF INVENTION: Encoded At The Self-Incompatibility Locus
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yahwak & Associates
; STREET: 25 Skytop Drive
; CITY: Trumbull
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06611
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07717,331F
; FILING DATE: June 19th 1991
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: George M. Yahwak
; REGISTRATION NUMBER: 26,824
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203)268-1951
; TELEFAX: (203)268-1951
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 857 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-717-331F-2

```

Db 136 DLSAKNLNLTLSNQSEVSLDLEFIPTNFSDKNPLKLPEDGSVVVANE---SYTVSVK 191
QY 236 VKYKAEDLMGTDSDLMFGYTOQSHWQIFNGKNSRPFVHDYQPEI-FLTQPYSDL-PW 293
Db 192 ATLQKLVLYSNADSVGISTAIPTVSLNGKTONDFSEFPKSNINFAKNYNNALNPF 251
QY 294 DGVNRIGMGAVHHS--NGESAK-----LSRSMNRAYLAGMEKMLTVMPRIWGRIEK 345
Db 252 EAQOYLVGQGFKNQKVNADVDKNDINNHETQFNVAKITA-----TLGKAFK 300
QY 346 ---EGSGOP-----DDNPILDY--GYGDV-----RFLYQL-- 373
Db 301 QGFCHKNQGLSLKVLKSLGSLNNEFKOLFVTPRGDGVSDLIQSSSQSSNKKTYQLLF 360
QY 374 ENKSN 379
Db 361 ENKTTI 366

RESULT 7

US-07-946-497-2
; Sequence 2, Application US/07946497
; Patent No. 5506119

GENERAL INFORMATION:

APPLICANT: HERRLICH, Peter
APPLICANT: PONTA, Helmut
APPLICANT: GUENTHERT, Ursula
APPLICANT: MATZKU, Siegfried
APPLICANT: WENZL, Achim
TITLE OF INVENTION: VARIANT C044 SURFACE PROTEINS, DNA
TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/946,497
FILING DATE: 19921109

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16915/145
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-946-497-2

Query Match

Best Local Similarity 3.8%; Score 89.5; DB 1; Length 503;
Matches 57; Conservative 32; Mismatches 96; Indels 125; Gaps 14;

QY 40 ENDLGDNEPLTIDVOSATQSASTDTANPLDEHPELYTTALEKTKMLINSALNODIMRL 99
Db 169 QEDIDASNITIDEVSS-----GSTLEKSPREGYIILHTDLP---SQPTGDRD----- 212

QY 100 ACYDTLVHGETPAVVIKTKRSI-----RIDETIMQTIKGRPOVLYOETTDPIFLMGNEKG 153
Db 213 ---DAFIEGSLTALITTPVWSAHTKONERTQWNPISHNPEVLQTTT----- 258
QY 154 MLTKKAKOLEYAKQOTPLSLSFDDLRNNTPL---WSSRPINPMVVLIFMHGKPNRS 209
Db 259 -----RNT-----DIDRNSTAHGEMWTQEPQPPF----- 283
QY 210 PNTPSHAKQFTPNFAPRLKFOVSVKRAEDLW-----GTDSDLMEGYTOQS 259
Db 284 -NNHEYODEEETH-----ATSTTADPNSTTEATQAKENF-----EN 322
QY 260 HWOIFNGKN-----SRPFVHDYQPEIFLTQPYSDLPW---DGKVRMIGMG 303
Db 323 EMQ---GKNPPTSEDSHVTGTTASAHNNHPSQRMTTQSQEDVSWTDFDPLISHPMGQG 379
QY 304 AVHNSGESA 313
Db 380 HOTESKGHSS 389

RESULT 8

US-08-483-322-2
; Sequence 2, Application US/08483322
; Patent No. 5760178

GENERAL INFORMATION:

APPLICANT: HERRLICH, Peter
APPLICANT: PONTA, Helmut
APPLICANT: GUENTHERT, Ursula
APPLICANT: MATZKU, Siegfried
APPLICANT: WENZL, Achim
TITLE OF INVENTION: VARIANT C044 SURFACE PROTEINS, DNA
TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,322
FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/946,497
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16915/145
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-483-322-2

Query Match

Best Local Similarity 3.8%; Score 89.5; DB 1; Length 503;
Matches 57; Conservative 32; Mismatches 96; Indels 125; Gaps 14;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:14:03 ; Search time 36.79 Seconds
(without alignments)
270.358 Million cell updates/sec

Title: US-09-787-083-2
Perfect score: 2360
Sequence: 1 MKVSLTSLTSLSCFAILA.....YNHEATSGVGLMLNDWGL 442

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/1aa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PCTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101.5	4.3	1726	2	US-08-609-049A-30
2	101.5	4.3	1726	4	US-09-170-996-30
3	100.5	4.3	1658	2	US-08-609-049A-13
4	100.5	4.3	1658	4	US-09-170-996-13
5	95.5	4.0	682	3	US-08-481-435-6
6	91.5	3.9	1024	4	US-09-091-117-5
7	89.5	3.8	503	1	US-07-946-497-2
8	89.5	3.8	503	1	US-08-483-322-2
9	89.5	3.8	503	2	US-08-478-882-2
10	89.5	3.8	857	1	US-07-717-331F-2
11	87	3.7	537	2	US-08-633-879C-2
12	87	3.7	985	5	PCT-US96-03916-6
13	87	3.7	985	5	PCT-US96-03916-66
14	86.5	3.7	666	4	US-08-961-083-2
15	86	3.6	781	1	US-08-373-134D-2
16	86	3.6	781	2	US-09-114-637-2
17	86	3.6	1627	1	US-07-665-792E-9
18	84	3.6	522	6	RE34606-6
19	84	3.6	816	1	US-07-731-157A-4
20	84	3.6	816	1	US-08-229-444B-2
21	84	3.6	816	2	US-08-541-780-4
22	83.5	3.5	774	3	US-08-902-632-2
23	83.5	3.5	774	3	US-09-073-354-1
24	83.5	3.5	774	3	US-08-656-005A-1
25	83.5	3.5	774	4	US-09-073-259-1
26	83.5	3.5	774	4	US-09-363-095-1
27	83.5	3.5	774	4	US-09-418-027-1

28	83	3.5	355	1	US-07-946-497-5	Sequence 5, Appl
29	83	3.5	355	1	US-08-483-322-5	Sequence 5, Appl
30	83	3.5	355	2	US-08-478-882-5	Sequence 5, Appl
31	82.5	3.5	535	2	US-08-564-972-1	Sequence 1, Appl
32	82.5	3.5	535	4	US-09-171-969-9	Sequence 9, Appl
33	82.5	3.5	657	4	US-09-306-593-2	Sequence 2, Appl
34	82.5	3.5	663	1	US-08-441-139-7	Sequence 7, Appl
35	82	3.5	334	2	US-08-359-850-4	Sequence 4, Appl
36	82	3.5	527	2	US-08-592-126-145	Sequence 145, App
37	82	3.5	527	2	US-08-687-080-48	Sequence 48, Appl
38	82	3.5	535	2	US-08-633-879C-4	Sequence 4, Appl
39	82	3.5	797	4	US-09-086-912-2	Sequence 2, Appl
40	82	3.5	1022	3	US-08-772-270A-2	Sequence 2, Appl
41	82	3.5	1088	3	US-08-633-768A-1	Sequence 1, Appl
42	81	3.4	320	2	US-08-245-511-4	Sequence 4, Appl
43	81	3.4	320	2	US-08-600-993A-4	Sequence 4, Appl
44	80.5	3.4	354	1	US-07-946-497-4	Sequence 4, Appl
45	80.5	3.4	354	1	US-08-483-322-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-609-049A-30
; Sequence 30, Application US/08609049A
; Patent No. 5948664
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; TITLE OF INVENTION: Chen, Yen-Wen
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,049A
; FILING DATE: 29-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-063700US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1726 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-609-049A-30

Query Match 4.3%; Score 101.5; DB 2; Length 1726;
Best Local Similarity 18.6%; Pred. No. 0.47;
Matches 98; Conservative 80; Mismatches 189; Indels 161; Gaps 21;
QY 1 MKVSL-----TTLTSLTSLSCFAILAIOQAQAVNPVAFVD-----EVRSENDLQD 46
DB 465 VKVSIIEGLQLPVTTCDSVSTVEIIMQALSWHDDLNQVDVGSYILKVCQEEVLQN 524

QY 47 NELPIDVQSATQSDTDANPLDEHEPELYTTALENKTMLINCNSALNODIMRLACYDTLV 106
Db 525 NHCGLSHEHICRQKWDI-----EIKQLLTLSAMCONLARTAEDD--- 565
QY 107 HGETPAVITKRSIRLDETIMQTIKGPQVIYQETDPIF----- 146
Db 566 --EAP-----VDLNKYLQIEKPYKEVMTVRHPVEELDSYHYQVELALQATENQHRV 615
QY 147 -----LMGNEKGMTKKDAKOLEYAA-----KQFTPLSLSFDDLDRNNTPLWS 188
Db 616 DOVIKAVRKICSDGVETPSVTEA-VKKLRAVNLPRKNSADVTSLSGSDTRKNTKGS 674
QY 189 SRPHNPMYVLP-----IFMHGKPNR-----SPNTPSHEAKOFTNFRAPELKQV 234
Db 675 LNPNPVOVSDMHLTRIYDILLRLHANSRCSGTGCPGRSRNIKEAWTATE-----QL 726
QY 235 SVKKAEDL---WCTSDLMFGYTOQSHWQIFNCKN-SRPF---RVHDYQPEIFLTQ--- 285
Db 727 QFTVYAAHGISSNMVSNYEKYILCSLSH---NGKDLFKPIQSKKVGTYKNFFYLKWD 782
QY 286 -----PV-YSDLPWDGKYRMIGMGAVHHSNGESAKLSRSWN-----RAYL 324
Db 783 ELIIFPIQISQLPLESVLHLTLFGVLNQSNGSSPSDNKQKRGPEALGVSLTLDFKRF 842
QY 325 MAGMEWKNTVMPIRWIRIFKEGSGQDPDNDIILDYGYGVDRELYOLENKNISGTVR 384
Db 843 TCG-----TKLLYLW-----TSSTNSIPGAIPKKSVMYERIVLQVDFPSPAFDIY 889
QY 385 YNPRSGKALQLDYVYPLGKIGSYFQIFQGYGQSLIDYNHEATSGV 432
Db 890 TSPQIDRNIIOQDKLETLESIDK-----KLLDIHRDSSFGL 927

RESULT 2

US-09-170-996-30
; Sequence 30, Application US/09170996
; Patent No. 6291220
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 6291220el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/170,996
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/609,049
; FILING DATE: 29-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-0637000S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1726 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-170-996-30

Query Match 4.3%; Score 101.5; DB 4; Length 1726;
Best Local Similarity 18.6%; Pred. No. 0.47;
Matches 98; Conservative 80; Mismatches 189; Indels 161; Gaps 21;

QY 1 MKVSLS-----TLTSLSCFAILAIQQAQVNPVAFVD-----EYRSENDLQGD 46
Db 465 VKVSIIEIGLQLPVFTTCDVSVSTEIIIMQALSWHDDLNQVDVGSYILKVCQGQEVQLN 524
QY 47 NELPIDVQSATQSDTDANPLDEHEPELYTTALENKTMLINCNSALNODIMRLACYDTLV 106
Db 525 NHCGLSHEHICRQKWDI-----EIKQLLTLSAMCONLARTAEDD--- 565
QY 107 HGETPAVITKRSIRLDETIMQTIKGPQVIYQETDPIF----- 146
Db 566 --EAP-----VDLNKYLQIEKPYKEVMTVRHPVEELDSYHYQVELALQATENQHRV 615
QY 147 -----LMGNEKGMTKKDAKOLEYAA-----KQFTPLSLSFDDLDRNNTPLWS 188
Db 616 DOVIKAVRKICSDGVETPSVTEA-VKKLRAVNLPRKNSADVTSLSGSDTRKNTKGS 674
QY 189 SRPHNPMYVLP-----IFMHGKPNR-----SPNTPSHEAKOFTNFRAPELKQV 234
Db 675 LNPNPVOVSDMHLTRIYDILLRLHANSRCSGTGCPGRSRNIKEAWTATE-----QL 726
QY 235 SVKKAEDL---WCTSDLMFGYTOQSHWQIFNCKN-SRPF---RVHDYQPEIFLTQ--- 285
Db 727 QFTVYAAHGISSNMVSNYEKYILCSLSH---NGKDLFKPIQSKKVGTYKNFFYLKWD 782
QY 286 -----PV-YSDLPWDGKYRMIGMGAVHHSNGESAKLSRSWN-----RAYL 324
Db 783 ELIIFPIQISQLPLESVLHLTLFGVLNQSNGSSPSDNKQKRGPEALGVSLTLDFKRF 842
QY 325 MAGMEWKNTVMPIRWIRIFKEGSGQDPDNDIILDYGYGVDRELYOLENKNISGTVR 384
Db 843 TCG-----TKLLYLW-----TSSTNSIPGAIPKKSVMYERIVLQVDFPSPAFDIY 889
QY 385 YNPRSGKALQLDYVYPLGKIGSYFQIFQGYGQSLIDYNHEATSGV 432
Db 890 TSPQIDRNIIOQDKLETLESIDK-----KLLDIHRDSSFGL 927

RESULT 3

US-08-609-049A-13
; Sequence 13, Application US/08609049A
; Patent No. 5948664
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 5948664el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,049A
; FILING DATE: 29-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-0637000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1658 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-609-049A-13

Query Match 4.3%; Score 100.5; DB 2; Length 1658;
Best Local Similarity 18.9%; Pred. No. 0.56;
Matches 101; Conservative 79; Mismatches 194; Indels 159; Gaps 23;

QY 5 LSTPLTSLSCFAILAIQ---QAQAVPNPVAFVDEYRS-----ENDLGQDNE 48
Db 381 LSPVTVQRMNCGENASVKVSIETGQLPVTFTCDVSSVTEIIMQALCWVHDDNQ--- 437

QY 49 LPIDVQS-----ATQSASTDTANPLDEHE--PELYTTALENKTMLNCSALNODIMRLAC 101
Db 438 --VDVGSYILKVCQGQEVLNQNHCLGSHHEIQNCRKWDTEIKLQLLTLSAMCQNLARTAE 495

QY 102 YDTLVHGETPAVTKRSIRLDETWTQIKGKPOVIYQETTDPIF----- 146
Db 496 DD-----EAP-----VDLNKYLQIEKPYKEVMIRHPVEELDSYHYQVELALQTEN 542

QY 147 -----LMGNEKGMLTKKDAKQLEYAA-----KQFTPLSLSFOLDNRN 183
Db 543 QHRAVDQVIKAVRKICSDLDGVETPSVTEA-VKKLRAVNLPRNKSAVTSLSGSDTRKN 601

QY 184 TPLWSSRPHPNPMVLP-----IFMHGKPNR-----SPNTPSHEAKQFTNPEFRAPE 229
Db 602 STKGSLNPNPQVQSMDDLTTAIYDILLRLHANSRSCSTGCPGRSRNIKEAWTATE----- 656

QY 230 LKFOVSVKKAABEDL---WGTDSDLWFGYTOQSHWQIFNGKN-SRPF---RVHDIQPEIF 282
Db 657 ---QLQFTVYAAHGISSNNVSVYKYLICSLSH-----NGKDLFKPIQSKKVGTYKNFFY 709

QY 283 LTQ-----PV-YSDLPWDGKVRMIGMGAHVHNSGESAKLSRSWN----- 320
Db 710 LIKWDELIIFPIQISQLPLESVLHLLTFLGVNLQSSGSDSNKQKRGPEALGKVSILTLD 769

QY 321 -RAYLMAGMEWKNLTVMPRIWGRIFKEGSGSQPDNDPILDYVYGVDVRFYOLENKSNI 379
Db 770 FKRLFTCG-----TKLLYLW-----TSSHTNSIPGAIPKKSYYMERIVLQVDFPSPA 816

QY 380 SCTVRYNPRSGKALQOLDYVYPLGKIGISGYFOIFOGYGSGLIDYNHEATSFVG 432
Db 817 FDIITYSPQIDRNIQODKLETLESIDKG-----KLLDIHRDSSFGL 859

RESULT 4
US-09-170-996-13
; Sequence 13, Application US/09170996
; Patent No. 6291220
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 6291220el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
```

```
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/170,996
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/609,049
; FILING DATE: 29-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-0637000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1658 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-170-996-13

Query Match 4.3%; Score 100.5; DB 4; Length 1658;
Best Local Similarity 18.9%; Pred. No. 0.56;
Matches 101; Conservative 79; Mismatches 194; Indels 159; Gaps 23;

QY 5 LSTPLTSLSCFAILAIQ---QAQAVPNPVAFVDEYRS-----ENDLGQDNE 48
Db 381 LSPVTVQRMNCGENASVKVSIETGQLPVTFTCDVSSVTEIIMQALCWVHDDNQ--- 437

QY 49 LPIDVQS-----ATQSASTDTANPLDEHE--PELYTTALENKTMLNCSALNODIMRLAC 101
Db 438 --VDVGSYILKVCQGQEVLNQNHCLGSHHEIQNCRKWDTEIKLQLLTLSAMCQNLARTAE 495

QY 102 YDTLVHGETPAVTKRSIRLDETWTQIKGKPOVIYQETTDPIF----- 146
Db 496 DD-----EAP-----VDLNKYLQIEKPYKEVMIRHPVEELDSYHYQVELALQTEN 542

QY 147 -----LMGNEKGMLTKKDAKQLEYAA-----KQFTPLSLSFOLDNRN 183
Db 543 QHRAVDQVIKAVRKICSDLDGVETPSVTEA-VKKLRAVNLPRNKSAVTSLSGSDTRKN 601

QY 184 TPLWSSRPHPNPMVLP-----IFMHGKPNR-----SPNTPSHEAKQFTNPEFRAPE 229
Db 602 STKGSLNPNPQVQSMDDLTTAIYDILLRLHANSRSCSTGCPGRSRNIKEAWTATE----- 656

QY 230 LKFOVSVKKAABEDL---WGTDSDLWFGYTOQSHWQIFNGKN-SRPF---RVHDIQPEIF 282
Db 657 ---QLQFTVYAAHGISSNNVSVYKYLICSLSH-----NGKDLFKPIQSKKVGTYKNFFY 709

QY 283 LTQ-----PV-YSDLPWDGKVRMIGMGAHVHNSGESAKLSRSWN----- 320
Db 710 LIKWDELIIFPIQISQLPLESVLHLLTFLGVNLQSSGSDSNKQKRGPEALGKVSILTLD 769

QY 321 -RAYLMAGMEWKNLTVMPRIWGRIFKEGSGSQPDNDPILDYVYGVDVRFYOLENKSNI 379
Db 770 FKRLFTCG-----TKLLYLW-----TSSHTNSIPGAIPKKSYYMERIVLQVDFPSPA 816

QY 380 SCTVRYNPRSGKALQOLDYVYPLGKIGISGYFOIFOGYGSGLIDYNHEATSFVG 432
Db 817 FDIITYSPQIDRNIQODKLETLESIDKG-----KLLDIHRDSSFGL 859

RESULT 5
US-08-481-435-6
```

; Sequence 6, Application US/08481435
; Patent No. 6027906
; GENERAL INFORMATION:
; APPLICANT: Balganes, Tanjore S
; APPLICANT: Town, Christine
; TITLE OF INVENTION: No. 6027906el Polypeptides
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481.435
; FILING DATE: 10-JUL-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IN 580/MAS/94
; FILING DATE: 01-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9404072-2
; FILING DATE: 24-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-151
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 682 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-481-435-6

Query Match 4.0%; Score 95.5; DB 3; Length 682;
Best Local Similarity 20.4%; Pred. No. 0.43;
Matches 92; Conservative 66; Mismatches 191; Indels 103; Gaps 24;
Qy 13 LSCFAILAIQQAQVNPVAFVDEVRSENDL-GQDNELPIDVQSATQS-----AST 62
Db 5 LSESKLVATTSSKIYDNKNQLIADGSERRVNAQANDIPTDLVKAIVSTEDHRRFDHRCI 64
Qy 63 DTANPLDEHEPELYTTALENTKMLNCSALNODIMRLACYDTLVHGETPAVTKTKRSIRL 122
Db 65 DTIRLGAFLRLQNSLQGG-----SALTQQLIKLVFSTSTSDQIS-----RK 110
Qy 123 DETIWTQTI---KGRPVQIYQETDPIELMGNEKMLIK-----KDAKQLE----- 164
Db 111 AQEAWLAIQLOKATKQELIYYINKVYMSNGNYGMQTAQNYKDKLNNLSLPOLALLA 170
Qy 165 ---YAAKQFTPLS-LSFOLDNRNTPLWSRPHNPVYLPFIEMHGKPNRSPNTPPSHEAKQF 220
Db 171 GMPQAPNQYDPSHPAAQDRRLVL---SEMKNOGYI---SAEQYKAVNTPITDGLQS 224
Qy 221 TPNFERAPE-----LKFOVS-VKVAABEDLMGTDSDLMFGYQTQ--QSH-WOIFNGKNRSP 271
Db 225 LKASNYPAYMDNLVKEVINQVEETGYNLLTTGMDVYTNVDQEAQKHLWDIYNTDEYVA 284
Qy 272 FRVHDYQPEIFLTQPVYSDLPWDGKVRMIGCAVHHSN-----GESAKLSRSNRYLM 325
Db 285 YPDELQ-----VASTIVDVS-NGKV-IAQLGARHQSSNVSGFINGQAVETNRDW----- 331

Qy 326 AGMEWKNTL-VMPRIWGRIFEGSGSQDDNPDILDIDYG-----YGDVRELYQ 372
Db 332 -GSTWKPIITDYAPALEYGVY-ESTATIVHDEP--YNYGCTNTPVYNWDRGTFGNTILQYA 387
Qy 373 LENKSNISGTVRYN-----PRSGKGAQLQDY 398
Db 388 LQQRNVPAVETLNKVGLENRAKTFNLGLGIDY 419
RESULT 6
US-09-091-117-5
; Sequence 5, Application US/09091117
; Patent No. 6171589
; GENERAL INFORMATION:
; APPLICANT: The University of Melbourne
; TITLE OF INVENTION: Mycoplasma Recombinant Polypeptides and
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GREENLEE, WINNER and SULLIVAN P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/091.117
; FILING DATE: 12 JUNE 1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU96/00803
; FILING DATE: 13-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PN7127
; FILING DATE: 13-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: WINNER, Ellen P.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: +1 303 499 8080
; TELEFAX: +1 303 499 8089
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1024 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORGANISM: Mycoplasma genitalium
US-09-091-117-5

Query Match 3.9%; Score 91.5; DB 4; Length 1024;
Best Local Similarity 20.2%; Pred. No. 2.2;
Matches 86; Conservative 61; Mismatches 172; Indels 107; Gaps 22;
Qy 1 MKVS-LSTLTLSISCFAILAIQQAQVNPVAFVDEVRSENDLGQDNELPT-DVQSATQ 58
Db 1 MKLSTITITICLSISGAFGTAI---ALPTTVALLKNHQOQNTKEQON--PIKDIRFGLN 54
Qy 59 SASDTPANPLDEHEPELYTTALENTKMLNCSALNODIMRLACYDTLVHGETPAVTKTKR 118
Db 55 NVQVPNTIPL--HQTVVEVT--NNKAIYDYKDAPOKFFL-----AKSALNNKL 98
Qy 119 SIRLDETWTQITKGPQVIYQETDPIELMGNEKMLTKDAKQLEYAAKO-FTPLSLSF 177
Db 99 QVEFDKFLRT-----GVINNALNADLKEWIDQITLFIQNQSF 135
Qy 178 DLDRNNTPLWSRPHNPVYLP-IPMHGKPNRSP-NTPSHEAKQETPNFERAPELKFOVS 235

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:17:03 ; Search time 77.9 Seconds
(without alignments)
829,941 Million cell updates/sec

Title: US-09-787-083-2
Perfect score: 2360
Sequence: 1 MKVSLTLTLILSCFAILA.....YNHEATSFVGLMLNDWMGL 442

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	814	34.5	382	2	Q9K0U7	Q9K0U7 neisseria m
2	814	34.5	409	2	Q9JUT21	Q9JUT21 neisseria m
3	372	15.8	306	2	Q9CL22	Q9CL22 pasteurella
4	367	15.6	329	2	Q9PMU8	Q9PMU8 campylobact
5	360.5	15.3	292	2	Q9SLD7	Q9SLD7 versinia ps
6	360	15.3	289	2	Q9L6N9	Q9L6N9 salmonella
7	356.5	15.1	292	2	Q9Z4N8	Q9Z4N8 enterobacte
8	342.5	14.5	297	2	Q9Z349	Q9Z349 campylobact
9	246.5	10.4	355	2	Q92541	Q92541 helicobacte
10	240.5	10.2	355	2	Q9ZLX5	Q9ZLX5 helicobacte
11	122	5.2	278	2	Q9XB53	Q9XB53 erwinia car
12	117	5.0	1686	4	O00443	O00443 homo sapien
13	112.5	4.8	602	11	Q63485	Q63485 rattus norv
14	110.5	4.7	821	2	Q59241	Q59241 bacillus sp
15	109	4.6	824	2	Q9F216	Q9F216 bacillus sp
16	107.5	4.6	901	5	Q18749	Q18749 caenorhabdi
17	105	4.4	435	2	Q9X9C0	Q9X9C0 streptococc
18	105	4.4	783	2	Q45554	Q45554 bacillus sp
19	103.5	4.4	1046	2	O84941	O84941 streptococc

20	103	4.4	798	4	Q9UN32	Q9UN32 homo sapien
21	103	4.4	798	4	Q9UBK2	Q9UBK2 homo sapien
22	102.5	4.3	3247	12	Q65553	Q65553 bovine herp
23	101.5	4.3	403	5	P91736	P91736 hydra magni
24	101.5	4.3	660	5	Q9KGX8	Q9KGX8 bacillus ha
25	101.5	4.3	739	5	Q9V9E6	Q9V9E6 drosophila
26	101.5	4.3	838	10	Q9C815	Q9C815 arabidopsis
27	100.5	4.3	1509	11	Q61194	Q61194 mus musculu
28	100.5	4.3	1658	11	Q61182	Q61182 mus musculu
29	99.5	4.2	788	10	Q9SCV5	Q9SCV5 arabidopsis
30	99	4.2	422	2	Q9RCZ6	Q9RCZ6 streptomyce
31	99	4.2	5005	2	Q9PP25	Q9PP25 ureaplasma
32	98.5	4.2	467	4	Q9UJMS	Q9UJMS homo sapien
33	98.5	4.2	467	4	Q9BTY2	Q9BTY2 homo sapien
34	98.5	4.2	797	2	Q9RDW6	Q9RDW6 lactobacilli
35	97.5	4.1	683	2	Q9A6U7	Q9A6U7 caulobacter
36	97	4.1	576	2	Q9L1I5	Q9L1I5 streptomyce
37	97	4.1	4307	5	Q9J319	Q9J319 caenorhabdi
38	96	4.1	871	4	Q9Y5C3	Q9Y5C3 homo sapien
39	96	4.1	938	4	Q9Y5F7	Q9Y5F7 homo sapien
40	95	4.0	690	5	O61142	O61142 plasmodium
41	95	4.0	719	2	O57114	O57114 streptococc
42	95	4.0	719	2	Q9RET8	Q9RET8 streptococc
43	95	4.0	791	10	O49137	O49137 arabidopsis
44	94.5	4.0	476	5	Q9BI60	Q9BI60 caenorhabdi
45	94.5	4.0	530	5	O45879	O45879 caenorhabdi

ALIGNMENTS

RESULT 1
Q9K0U7 PRELIMINARY; PRT; 382 AA.
AC Q9K0U7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PHOSPHOLIPASE A1, PUTATIVE.
OS NMB0464.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B., Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J., Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;
RA "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58."
RT Science 287:1809-1815(2000).
RL EMBL: AE002403; AAF40901.1; -.
DR TIGR: NMB0464; -.
DR InterPro: IPR003187; PLAL.
DR Pfam: PF02253; PLAL; 1.
KW Complete proteome.
SQ SEQUENCE 382 AA; 42714 MW; B468A802F062E836 CRC64;

Query Match 34.5%; Score 814; DB 2; Length 382;
Best Local Similarity 44.4%; Pred. No. 1e-62;
Matches 162; Conservative 68; Mismatches 111; Indels 24; Gaps 8;

QY 87 INCSALNDIMRLACYDTLVHGETPAVI-----KTKRSIRLDETINQTI-KGKPVQIYQE 140
DB 33 LOCALTDNTRLYACIDYRFAAQLPSSAQGEQSKAVLNLTETVRSSLDKGEAVIVVEK 92

[illegible]

```

Qy 345 KEGSQDDNDPILDYGYGDRFLYOLENKSNSGTVRNPRSGKALOLDYVYPLGK 404
Db 200 -----GSTDDNDPITYKMGYQLKGYHL-GEAVLSAKQYNWNTGYGAEVGLSPYTK 253

Qy 405 GISGYFQIFQGYGOSLDYDYNHEATSGVGLMLND 438
Db 254 HVRLYTQVSYGSGESLDYDYNFQTRGVGVMLND 287

RESULT 7
Q924N8 PRELIMINARY; PRT; 292 AA.
AC Q924N8;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE OUTER MEMBRANE PHOSPHOLIPASE A PRECURSOR (EC 3.1.1.32).
GN PLDA.
OS Enterobacter agglomerans.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94131966; PubMed=8300539;
RA Brok R.G., Brinkman E., van Bostel R., Bekkers A.C., Verheij H.M.,
RA Tommassen J.;
RT "Molecular characterization of enterobacterial pida genes encoding
RT outer membrane phospholipase A.,"
RL J. Bacteriol. 176:861-870(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120145; PubMed=9921577;
RA Brok R.G., Boots A.P., Dekker N., Verheij H.M., Tommassen J.;
RT "Sequence comparison of outer membrane phospholipases A: implications
RT for structure and for the catalytic mechanism.,"
RL Res. Microbiol. 149:703-710(1998).
DR EMBL; AF034414; AAD03498.1; -.
DR InterPro; IPR003187; PLA1.
DR Pfam; PF02253; PLA1; 1
DR PRINTS; PR01486; PHPLIPASEAL.
KW Signal; Hydrolase.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 292 OUTER MEMBRANE PHOSPHOLIPASE A.
SQ SEQUENCE 292 AA; 33719 MW; B75516D093B2BEEA CRC64;

Query Match 15.1%; Score 356.5; DB 2; Length 292;
Best Local Similarity 40.1%; Pred. No. 5.2e-23;
Matches 85; Conservative 31; Mismatches 87; Indels 9; Gaps 5;

Qy 229 ELKFQVSKVKAEDLWCTGDSLWFGYTOQSHWQIFNCKNSRPRFRVHDYQPEIFLTQPV- 287
Db 86 EVKQLSGFFPIWRCIAGNSLLGASYTORSMOANSDESPPRETNYEPOIFLAWTD 145

Qy 288 YSCLPQWCKVRMIGMGAHVHSGESAKLSRWNRAYLMAGMEWKNLTVMPIRWGRIFKEG 347
Db 146 YELAGW--TFREVEFGYNHQSNRGPDSRSMDRYVTRLMQARGNLEIDLPWYRIPESD 203

Qy 348 SGSPDDNDPILDYGYGDRFLYQL-ENKSNISGTVRNPRSGKALOLDYVYPLGKI 406
Db 204 S---KDDNPDINKYGYRLRVGVALGESVFSIDG--RYNNTGYGGAEMGWSYPIKRV 258

Qy 407 SGYFQIFQGYGOSLDYDYNHEATSGVGLMLND 438
Db 259 RFTYQVSGYSGESLDYDYNFQTRGVGVMLND 290

RESULT 8
O32349 PRELIMINARY; PRT; 297 AA.
ID O32349
AC O32349;

```

```

DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE PHOSPHOLIPASE A.
GN PLDA.
OS Campylobacter coli.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=195;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=UA585;
RX MEDLINE=97230284; PubMed=9119448;
RA Grant K.A., Belandia I., Dekker N., Richardson P.T., Park S.F.;
RT "Molecular characterization of pida, the structural gene for a
RT phospholipase A from Campylobacter coli, and its contribution to cell-
RL associated hemolysis.,"
RL Infect. Immun. 65:1172-1180(1997).
DR EMBL; Y11031; CAA71915.1; -.
DR InterPro; IPR003187; PLA1.
DR Pfam; PF02253; PLA1; 1
DR PRINTS; PR01486; PHPLIPASEAL.
SQ SEQUENCE 297 AA; 34998 MW; 04B54A7BCA3764CE CRC64;

Query Match 14.5%; Score 342.5; DB 2; Length 297;
Best Local Similarity 32.4%; Pred. No. 8.8e-22;
Matches 97; Conservative 37; Mismatches 108; Indels 57; Gaps 11;

Qy 178 DLDRNPTPLWSSRPHNMYVLPIFMHGKPNRSPNT-----P 213
Db 16 DLKENNASLLSRKHETQ-----NTQKTPSTKEDFSRIALANYLGENSESFNPLGIS 66

Qy 214 SHEAKQFTP-----NEFRAPELKFOVSKVKAEDLWCTGDSLWFGYTOQSHWQIF 264
Db 67 SYKNMYLFPFAYSGSLGGENRKTEMKFLSIKKRLFDLLGLGKGYVYGTQTSWQ-- 124

Qy 265 NGKNSRPRFRVHDYQPEIFLTQPV-YSDLPWCKVRMIGMGAHVHSGESAK--LSRSWNR 321
Db 125 NYKHSSPERETNYQPEFFVDIPLHEDYKFLNLR--VGILHESNGKGDENLESRSNR 181

Qy 322 AYLMAWMEWKNLTVMPIRWGRIFKEGSGQDDNDPILDYGYGDRFLYOLENKSNSIG 381
Db 182 IYASGVFLYQRFLEFVPIRWYRI---PENSEDDNDNPEITHYMGNFIDIN-IGSLGNDYFINL 237

Qy 382 TVRYNP--RSCKGKALQDYVYPL-CKGIGSGYFQIFQGYGOSLDYDYNHEATSGVGLMLN 437
Db 238 MLRNLDLFDHDKGAVQVDIGDIFDNGIYWIYQIFNGYGGSLDIDYKRLQLRLSTAFLLS 296

RESULT 9
O25241 PRELIMINARY; PRT; 355 AA.
ID O25241
AC O25241;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE PHOSPHOLIPASE A1 PRECURSOR (DR-PHOSPHOLIPASE A).
GN HP0499
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,

```

Best local similarity 25.19; Recd. no. 2.06 15;
Matches 82; Conservative 41; Mismatches 117; Indels 87; Gaps 11;

Db 69 KKYLNMDYLGTYFLPFYHSF-----TPIQWYHPNINP----- 102

Db	103	---	YQRNEF	----	KFQISFRVPVFRHILWT	KGTL	YLAY	TQT	DWF	QI	YND	PQ	SAP	MR	MM	N	154
Db	103	---	YQRNEF	----	KFQISFRVPVFRHILWT	KGTL	YLAY	TQT	DWF	QI	YND	PQ	SAP	MR	MM	N	154

Db 155 FMPELIYVYPI-NFKPFGKIGNFSEIWIGWQHISNGVGGAQCYQPFNK----EGNPENQ 209

Db 210 FPGQPIVKDYNQKDVRWGGCRSVSAGQRPVRLVWEKGLKIMVAYWPYPYDQSNPN 269

Db 270 LIDYMGYNAKIDYRRGRHFFELQLYDIFTQYWRD--RWHGAFLGYTYRINPFVGIYA 327

Db 328 QWENGCGDGLYEYDVFSNRIGVGIRLN 354

Q9ZLX5

DT 01-MAY-1999 (TREMBLrel. -10, Last sequence update)

OS Helicobacter pylori J99 (Campylobacter pylori J99).

RN [1]

RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,

RT "Genomic sequence comparison of two unrelated isolates of the human

DR InterPro; IPR003187; PLA1.

Query Match 5.2%; Score 122; DB 2; Length 278;
Best Local Similarity 21.3%; Pred. No. 0.013;
Matches 73; Conservative 53; Mismatches 128; Indels 88; Gaps 16;

6	TKQLISSILILVL-LLSLPFTALAAEGNTREDNFKLLGNDNVKVRSEAGALQLEQVDGQ	64
66	NPL-DEHEPELY-----TTALENKTMLINGSAL-----NODIMKLACY-DTLVHGTEP	111
65	MTLVDQHGKEIKQLRGMSTHGLQWPEITLNDNAYKALANDWESNMIRLAMYVGENGYASNP	124
112	AVIKTKRSIRLDETI-----WQT-----IKGKPOQVIY	138
125	ELTKSRVIKGIDIALENDMTYIVDWHVHAPGRDPRDPVYAGADEFRDIAALYNPNPHIYY	184
139	QETTPD-----IFLMGNEKGLTKKDAKQLEYAAKQFTPLSLSPDLDRN-----NTPLW	187
185	ELANEPSSNNGGAGIPNNEEGNAVK-----EYADPVEIMLRDSGNADDNIIVGSPNW	239
188	SSRP-----HNPWYVLPF--MHGKPNBS--PNTPSHAKQFTNPEFRAPELKFQ	233
240	SQRPDLAADNPIDDHTTMYTVHYTGSAASTESYPETPNBSRGNNVMSNTRYA-----LE	295
234	VSVKVAEADLWGT-----DSDLWFGYTOQ-----SHWQIFNGKNS-----RP	271
296	NGVAVFATE--WCTSQANGDGGPYFDEADWIEFLNENNISWANWSLTN-KNEVSGAFTP	352
272	FRVH-----DYQPEIFLTQPYISDLPW---D	294
353	FELGKSNAATSLDPGQDVWVPEELSLSGEYVRARIKGVNYEP---IDRTKYTKVLDFND	409
295	GKVRIMCGAVHHSGRS-----AKLSRSNRWAYLMAGMEKNLNTVMPRIWG	341
410	GTKGCGFV-----NGDSPVEDVVIENEAGALKLSGLDASNDYSEGNYHANARLSADGWG	463
342	R	342
464	K	464

RESULT 15
O9F216

```

ID Q9F216 PRELIMINARY; PRT; 824 AA.
AC Q9F216;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE CELLULOSE.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OC NCBI_TaxID=1409;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KSM-S237;
RX MEDLINE=2036886; PubMed=11193993;
RA Hakamada Y., Hatada Y., Koike K., Yoshimatsu T., Kawai S.,
RA Kobayashi T., Ito S.;
RT "Reduced amino acid sequence and possible catalytic residues of a
RT thermotstable, alkaline cellulase from an alkaliphilic Bacillus
RT strain.";
RL Biosci. Biotechnol. Biochem. 64:2281-2289(2000).
RL EMBL: AB018420; BAB19360.1; -.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR Pfam: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWNW.1.
SQ SEQUENCE 824 AA; 91564 MW; 65FA940FE1D729B9 CRC64;

```

Query Match	4.68;	Score 109;	DB 2;	Length 824;
Best Local Similarity	20.06;	Pred. No. 0.85;	159;	Indels 166;
Matches 96;	Conservative			Gaps 28;
Qy	9	TLTSLSCFAILAI	OOAQPVPVAFVDEVRSEND---LGDQNELPIDVOSATQSA	TDTA 65
		: :	: :	
Db	7	TQLISLTILILV	-LLSLFPAALAAEGNTRDNEFKHLLGNDGNVRPSPAGALQ	QVEVDGQ 65
Qy	66	NPL-DEHEPELY	-----TTALENTMTLINC	SAL-----NQDIMRLACYDTLVHGE--- 109

Db	66	MTLVDQHGKIQURGMSTHLOWFPBILNDNAYKALSNDDWDSNMIRLAWY----	VGNGY	121
Qy	110	--TPAVIKTK-----RSIRLDETI-----WQT-----	IKGKP	134
Db	122	ATNPELIKORVIDGIELATENDMYIVDVHVAHGGDPPVYAGAKDFRETAALYPNP	181	
Qy	135	QVIYQETTOP-----IFLMGNEKGMLTKKDAKOLEYAAKQFTPLSLSFOLDLRN-----	N	183
Db	182	HIYELANEPSSNNNGCAGIPNNEGWKAVK-----EYADPVLVEMRLKSGNADDMIIVG	236	
Qy	184	TPLWSSRP-----HNPYVLPVIFMHGKPNRSPTPSHEAKQFTPNFRAPELK-----	231	
Db	237	SPNKSQRPDLAADNPIDDHHTMYTHVFHVGSHAASTESYPSE-----TPNSERGNVMSNT	291	
Qy	232	---FQVSVKVKAAEDLWGT-----DSDLWFGYTOQ-----SHWQIFNGKNS-----	269	
Db	292	RYALENGVAVFATE--WGT'SQASGGDGGPYFDEADVWIEFLNENNTISWANWSLTN-KNEVS	348	
Qy	270	---RPRFV-----HDYQP-EIPLT-----QPV-----YSDLPW--	293	
Db	349	GAFTPFELGKSNATNLDPGDHWAPEELSLGSEYVRARIKGVNVEPDRTKYTKVLWDF	408	
Qy	294	-DGKVRMIGMGAHVHHSNGESAKLSRWSNRAYLMAGNE-----	-WKNLTVPMPRI	342
Db	409	NDGTQK--GFGVNSDSPNKELIAVDNENNTLKVSGLDVSDNYSDGNFWANARLSANGWK	466	

Search completed: November 30, 2001, 14:26:25
Job time: 562 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:17:53 ; Search time 28.02 Seconds
(without alignments)
578.367 Million cell updates/sec

Title: US-09-787-083-2

Perfect score: 2360
Sequence: 1 MKVSLSTLTLILSCFAILA.....YNHEATSGVGLMLNDWMGL 442

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	373	15.8	286	1 PAL_KLEPN	P37446 klebsiella
2	358	15.2	289	1 PAL_SALTY	P37442 salmonella
3	357	15.1	289	1 PAL_ECOLI	P00631 escherichia
4	321.5	13.6	289	1 PAL_PROVU	P37447 proteus vul
5	110.5	4.7	800	1 GUN_BACS1	P06564 bacillus sp
6	98.5	4.2	810	1 ANDM_YEAST	P15274 saccharomyc
7	96	4.1	888	1 YJH0_YEAST	P40361 saccharomyc
8	94.5	4.0	397	1 GUN_PAEPO	P23548 paenibacilli
9	93.5	4.0	1478	1 BCK1_YEAST	Q01389 saccharomyc
10	93	3.9	716	1 BAC2_MOUSE	P97303 mus musculus
11	92.5	3.9	849	1 SRK6_BRAOL	Q09092 brassica ol
12	92	3.9	324	1 YL70_ARCFU	O28112 archaeoglob
13	92	3.9	1158	1 RL14_HUMAN	P48552 homo sapien
14	92	3.9	1788	1 YP72_CAEEL	Q09221 caenorhabdi
15	91.5	3.9	467	1 INVO_MOUSE	P48997 mus musculus
16	91.5	3.9	486	1 VG14_BPM15	Q05220 mycobacteri
17	91.5	3.9	1024	1 Y075_MYCGE	P47321 mycoplasma
18	90.5	3.8	379	1 YJHT_HAEIN	P44544 haemophilus
19	90	3.8	496	1 CATA_DICDI	O77229 dictyosteli
20	90	3.8	719	1 PBPA_STRPN	Q04707 streptococ
21	89.5	3.8	503	1 CD44_RAT	P26051 rattus norv
22	89.5	3.8	825	1 GUN3_BACSA	P19570 bacillus sp
23	89	3.8	521	1 NPPE_BACAM	P06832 bacillus am
24	88.5	3.8	1379	1 MET_MOUSE	P16056 mus musculus
25	88	3.7	1131	1 PMAL_DUNBI	P54211 dunaliella
26	87.5	3.7	353	1 DCUP_BACSU	P32395 bacillus su
27	87.5	3.7	478	1 DHGB_ACICA	P13650 acinetobact
28	87.5	3.7	669	1 AMY_ALTHA	P29957 alteromonas
29	87.5	3.7	794	1 SEIL_HUMAN	Q9ubv2 homo sapien
30	87.5	3.7	969	1 SAGB_STRSL	Q55242 streptococ
31	87	3.7	483	1 MURE_CHLTR	O84271 chlamydia t
32	87	3.7	537	1 P4H2_MOUSE	Q60716 mus musculus
33	87	3.7	560	1 DTXH_CORBE	P00589 corynebphage

34 87 3.7 842 1 LPFC_SALTY P43662 salmonella
35 87 3.7 1115 1 DP3A_BACSU O34623 bacillus su
36 87 3.7 1474 1 A2MG_HUMAN P01023 homo sapien
37 86.5 3.7 1374 1 YC9A_SCHPO Q09884 schizosacch
38 86 3.6 491 1 TY3H_PHRASP P11982 phasianidae
39 86 3.6 666 1 PD14_RAT O88807 rattus norv
40 86 3.6 828 1 BGAL_BRAOL P49676 brassica ol
41 86 3.6 1627 1 ADP1_MYCPN P11311 mycoplasma
42 85.5 3.6 742 1 CD44_HUMAN P16070 h cd44 anti
43 85.5 3.6 790 1 SEIL_MOUSE Q92296 mus musculus
44 85 3.6 417 1 DHMH_PARDE P29894 paracoccus
45 85 3.6 517 1 EAS_DROME P54352 drosophila

ALIGNMENTS

RESULT 1
ID PAL_KLEPN STANDARD; PRT; 286 AA.
AC P37446;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DETERGENT-RESISTANT
DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
GN PLDA.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94131966; PubMed=8300539;
RA Brok R.G.P.M., Brinkman E., van Bostel R., Bekkers A.C.A.P.,
RA Verheij H.M., Tomassen J.;
RT "Molecular characterization of enterobacterial plda genes encoding
RT outer membrane phospholipase A.";
RL J. Bacteriol. 176:861-870(1994).
CC -!- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
CC -!- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -!- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -!- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
CC LOCATED THERE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X76901; CAA54223.1; -.
CC DR PIR; B36971; B36971.
CC DR PIR; S40129; S40129.
CC DR InterPro; IPR003187; PLAL.
CC DR Pfam; PF02253; PLAL; 1.
CC KW Hydrolase; Lipid degradation; Outer membrane; Signal; Calcium.
CC FT SIGNAL 1 20
CC FT CHAIN 21 286 PHOSPHOLIPASE A1.
CC FT ACT_SITE 161 161 BY SIMILARITY.
CC SQ SEQUENCE 286 AA; 32544 MW; 3E39F863085108A3 CRC64;

Query Match 15.8%; Score 373; DB 1; Length 286;
Best Local Similarity 42.1%; Pred. No. 8.1e-23;
Matches 90; Conservative 27; Mismatches 87; Indels 10; Gaps 5;

RL Nature 409:529-533(2001).
RP [5]
RN SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / RIMD 0509952;
RA Hayashi T., Makino K., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Lida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP SEQUENCE OF 30-289 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=85003590; PubMed=6383820;
RA de Gens P., Verheij H.M., Riegman N.H., Hoekstra W.P.M., de Haas G.H.;
RT "The pro- and mature forms of the *E. coli* K-12 outer membrane
RT phospholipase A are identical.";
RL EMBO J. 3:1799-1802(1984).
RN [7]
RP SEQUENCE OF 174-289 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=87115164; PubMed=3027506;
RA Irino N., Nakayama K., Nakayama H.;
RT "The *recQ* gene of *Escherichia coli* K12: primary structure and
RT evidence for SOS regulation.";
RL Mol. Gen. Genet. 205:298-304(1986).
RN [8]
RP MUTAGENESIS OF SER-172.
RX MEDLINE=94131966; PubMed=8300539;
RA Brok R.G.P.M., Brinkman E., van Bostel R., Bekkers A.C.A.P.,
RA Verheij H.M., Tommassen J.;
RT "Molecular characterization of enterobacterial plda genes encoding
RT outer membrane phospholipase A";
RL J. Bacteriol. 176:861-870(1994).
RN [9]
RP ACTIVE SITE SER-164, AND PARTIAL SEQUENCE.
RX MEDLINE=91249806; PubMed=2040286;
RA Horrevorts A.J.G., Verheij H.M., de Haas G.H.;
RT "Inactivation of *Escherichia coli* outer-membrane phospholipase A by
RT the affinity label hexadecanesulfonyl fluoride. Evidence for an
RT active-site serine.";
RL Eur. J. Biochem. 198:247-253(1991).
CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
CC REQUIRED FOR EFFICIENT SECRETION OF BACTERIOCINS, SEEMS TO BE
CC DORMANT IN NORMAL GROWING CELLS.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
CC LOCATED THERE.
CC -1- INDUCTION: BY MEMBRANE DAMAGING, FOR EXAMPLE, BY PHAGE-INDUCED
CC LYSIS OR TEMPERATURE SHOCK.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X02143; CAA26081.1; -;
DR EMBL; M87049; AAA67617.1; -;
DR EMBL; AE000458; AAC76824.1; -;
DR EMBL; AE005613; AAG59017.1; -;
DR EMBL; AP002567; BAB38174.1; -;
DR EMBL; M30198; AAA24516.1; -;
PIR; A00771; PSECA.

DR PIR; A22133; PSECAL.
DR PIR; S30711; S30711.
DR EcoGene; EGI0738; plga.
DR InterPro; IPR003187; PLAL.
DR Pfam; PF02253; PLAL; 1.
KW Hydrolase; Lipid degradation; Outer membrane; Signal; Calcium;
FT Complete proteome.
FT SIGNAL 1 20 PHOSPHOLIPASE A1.
FT CHAIN 21 289
FT ACT_SITE 164 164
FT MUTAGEN 172 172
FT CONFLICT 14 15 S->F: INACTIVE PROTEIN.
FT CONFLICT 30 33 LP -> FA (IN REF. 2).
FT SEQUENCE 289 AA; 33163 MW; A688AD32AA60F218 CRC64;
Query Match 15.1%; Score 357; DB 1; Length 289;
Best Local Similarity 39.3%; Pred. No. 1.6e-21;
Matches 84; Conservative 29; Mismatches 91; Indels 10; Gaps 4;
QY 226 RAPELKFOVSKVKAADLWGTDSLWFGYTCQSHWQIENGKNSRPRFRVHDYQPEIFLTQ 285
DB 83 RKDEVKFOLSLAFPLWRGILGPNVSLGASYTKSWQLSNSESSPRTNYPQLPLGF 142
QY 286 PV-YSDLPMDGKVRMIGMGAVHSHNGESAKLSRSNRAYLMAGMEKKNLTVMPIRNGRIF 344
DB 143 ATDYREFAGW--TLRDVEMGYNHDSNGRSDPTSRNRLYTRLMAENGWLVKPMYVW- 199
QY 345 KEGSGSQPDNDPILDYGYGVDFRFLYQLENKSNISGTVRYNPRSGKALQLODYVPLGK 404
DB 200 -----GNTDDNPDTITMGYKIGYHL-GDAVLAKGOYNWNTGYGGAELGLSYPTIK 253
QY 405 GISGYQIFOGYQSLIDYNHEATSFVGLMLND 438
DB 254 HVRLYQVSGYGESLIDYNFNQTRVGVGVMLND 287
RESULT 4
PAL_PROVU STANDARD; PRT; 289 AA.
AC P37447;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DETERGENT-RESISTANT
DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
GN PLDA.
OS Proteus vulgaris.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Proteus.
OX NCBI_TaxID=585;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94131966; PubMed=8300539;
RA Brok R.G.P.M., Brinkman E., van Bostel R., Bekkers A.C.A.P.,
RA Verheij H.M., Tommassen J.;
RT "Molecular characterization of enterobacterial plda genes encoding
RT outer membrane phospholipase A";
RL J. Bacteriol. 176:861-870(1994).
CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
CC LOCATED THERE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X02143; CAA26081.1; -;
DR EMBL; M87049; AAA67617.1; -;
DR EMBL; AE000458; AAC76824.1; -;
DR EMBL; AE005613; AAG59017.1; -;
DR EMBL; AP002567; BAB38174.1; -;
DR EMBL; M30198; AAA24516.1; -;
PIR; A00771; PSECA.

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X76902; CAA54224.1; -;
DR PIR; C36971; C36971;
DR PIR; S40130; S40130;
DR InterPro; IPR003187; PLA1.
DR Pfam; PF02253; PLA1; 1.
KW Hydrolase; Lipid degradation; Outer membrane; Signal; Calcium.
FT SIGNAL 20 BY SIMILARITY.
FT CHAIN 21 289 PHOSPHOLIPASE A1.
FT ACT_SITE 164 164 BY SIMILARITY.
SQ SEQUENCE 289 AA; 32944 MW; D75516CFFB406997 CRC64;

Query Match 13.6%; Score 321.5; DB 1; Length 289;
Best Local Similarity 33.3%; Pred. No. 1.le-18;
Matches 87; Conservative 42; Mismatches 115; Indels 17; Gaps 7;

QY 179 LDRNNTPLWSSRPHPMYPVLPFMHKGKPNRSPNTPSHEAKQFTPNFRAPELKFQVSVKV 238
DB 43 LOEHNP-FTLYPYSNLLIY-----TSLNKKAIKESYNWSDNA-NKDEVKFLSLAF 95
QY 239 KAAEDLWGTDSLWFGYTOQSHQIFNGKNSRPFVRVHDYQPEIFLTQPV-YSDLPWQGV 297
DB 96 PLWRGILGNSLGSYTORSMWLSNTGESAPFRETYEPQLFLGFATDYSVGW--TL 153
QY 298 RMIGMGAVHNSGESAKLSRSNRAYLMAGMEKWLTVMPRTWGRIFKGGSGQDDNDP 357
DB 154 RDAEFGYHQNSGDPSTSRNRLYSRLMAQNGNLVEVKPVI-----GDTSDNKN 207
QY 358 IDYYGYGDVRLYLENKSNISSVRYNPRSGKALQLDYVPLGKISGYFOIFQGVG 417
DB 208 ITKMGYQKIGYQL-GEVLASAGQYNWNTGYGAELGVSYPITKHFRFTQVYSGTG 266
QY 418 QSLIDYNHEATSFYGLMLND 438
DB 267 ESLIDYDFNQTRVGMVGLND 287

RESULT 5
GUN_BACSL
ID GUN_BACSL STANDARD; PRT; 800 AA.
AC P06564;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (ALKALINE CELLULOSE).
OS Bacillus sp. (strain 1139).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1411;
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE=87085443; PubMed=3098909;
RA Fukumori F., Kudo T., Narahashi Y., Horikoshi K.;
RT "Molecular cloning and nucleotide sequence of the alkaline cellulase gene from the alkalophilic Bacillus sp. strain 1139.";
RL J. Gen. Microbiol. 132:2329-2335(1986).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.
CC -1- MISCELLANEOUS: ALKALOPHILIC BACILLUS SP., STRAIN 1139, PRODUCES ONE ALKALINE CELLULOSE (PH OPTIMUM 9.0). THIS STRAIN IS NOT A TRUE CELLULOTIC MICRO-ORGANISM BECAUSE THE ENZYME IS UNABLE TO HYDROLYSE NATIVE CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M15743; AAA22305.1; -;
DR EMBL; D00066; BAA00045.1; -;
DR PIR; A29003; A29003.
DR InterPro; IPR001547; Glyco_hydro_F5.
DR Pfam; PF0150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL-HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 800 ENDOGLUCANASE.
FT ACT_SITE 190 190 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 305 305 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 800 AA; 88602 MW; 7CCA4D7B6DAD55CF CRC64;

Query Match 4.7%; Score 110.5; DB 1; Length 800;
Best Local Similarity 20.0%; Pred. No. 0.42;
Matches 96; Conservative 53; Mismatches 163; Indels 169; Gaps 25;

QY 9 TLSILSCFAILAIQQAQVNPVAFVDEVRSEND---LGQDNELPIDVQSASTDTA 65
DB 7 TKQLISSILILVL-LLSLFPPTALAAEGNTREDNFKHLLGNDNVKRPSEAGALQLQEVQ 65
QY 66 NPL-DEHEPELY-----TTALENKTMLNCSAL-----NQDINLACY-DTLVHGETP 111
DB 66 MTLVDPQHGEKIQLRGMSTHGLQWFFELINDNAYKALANDWESNMIRLAMYGVGYSNP 125
QY 112 AVIKTKRSIRLDETI-----WQT-----IKGKPOVY 138
DB 126 ELIKSRVIGKIDLAENDMYVIVDHWVHAPDRDPVYAGAEFFRDTAALYNNPHIY 185
QY 139 QETDP-----IFLMGNEKGLTKDAKQLEYAAKQFTPLSLSFDLDRN-----NTPLM 187
DB 186 ELANEPSSNNNGGAGIPNNEGNAVK-----EYADPIVEMLRDSGNADDNIIIVGSPNW 240
QY 188 SSRP-----HNPXVLPPIF--MHGKPNRS--PWTPSHEAKQFTPNFRAPELKFQ 233
DB 241 SQRPDLAANDPIDDHTTYTVHFTYTGSHAASTESYPPETPNSEKGNVMSNTRYA----LE 296
QY 234 VSVKVKAAEDLWGT-----DSDLWFGYTOQ-----SHWQIFNGKNS-----RP 271
DB 297 NCVAVFATE--WGTQANGCGGPFDEADVWIEFLNENNISWANNSLTN-KNEVSGATP 353
QY 272 FRVH-----DYQPEIFLTQPVYSDLPW---D 294
DB 354 FELGKSNATSLDPGDPQVWVPEELSLSGEYVRARIKGVNYEP---IDRTKYTKVLWDFND 410
QY 295 GKVRMIGMGAVHNSGES-----AKLSRSNRAYLMAGMEKWLTVMPRTW 341
DB 411 GTKQGFV-----NGDSPVEDVVIENEAGALKSLGDASNDVSEGNVWANRLSADGM 464
QY 342 R 342
DB 465 K 465

RESULT 6
AMDM_YEAST
ID AMDM_YEAST STANDARD; PRT; 810 AA.
AC P15274;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE AMP DEAMINASE (EC 3.5.4.6) (MYOADENYLATE DEAMINASE).
GN AMD1 OR AMD OR YKL035C.
OS Saccharomyces cerevisiae (Baker's yeast).
CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

[illegible]

Db 237 CTKNVYSAPSHGTSFASTFSDSPGNSLKGPLGPMGQIKSEPPSEETEEITLCLSGDE 296
 Qy 125 TIWQTIKGPQVIQYETDPI-----FLMNEKGMLTK 157
 Db 297 T---DIKORPGDVENDKQKPSARTPTGTGACILDRSRSVSSPCLSLFSLFITKGV--- 350
 Qy 158 KDAKOLEYAAKQFTPLSLF-----DLDRNNTPLWSS--RPHNPMYVLPFIMHG 204
 Db 351 -ESTGLPSTSQ--PLVRSSACPENKGISQGLKTDYPLAGNYGPHVQKQDVSNFAMG 407
 Qy 205 KPNRSP-----NTSHEAKQFTPEFRAPELK-----FQSVKVKAAEDLNGTSD 250
 Db 408 SPLRPGPETICEFSSPCSQARELAYEHOEPGLMGDMQVNOVRPQIK-CQSYGTNS 466
 Qy 251 LWFQYTOQSHWQIFNGKNSRPRVRHYDPEIFLTPV--YSDLPWDGKVRMIGMCAVHHS 308
 Db 467 DESG-----SFSEADESCVPDQGEVKLPFPVQITDLPNDPQMIMK---HKL 515
 Qy 309 NGESAKL-----SRSMNR 321
 Db 516 TSEQLEFIHDIRRSRKNR 533

RESULT 11
 SRK6_BRAOL STANDARD; PRT; 849 AA.
 AC Q09092;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE PUTATIVE SERINE/THREONINE KINASE RECEPTOR PRECURSOR (EC 2.7.1.37)
 DE (S-RECEPTOR KINASE) (SRK).
 GN SRK6.
 OS Brassica oleracea (Cauliflower).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3712;
 RN SEQUENCE FROM N.A.
 RC STRAIN=CV. S656; TISSUE=Stigma;
 RX MEDLINE=92020942; PubMed=1681543;
 RA Stein J.C., Howlett B., Boyes D.C., Nasrallah M.E.;
 RT "Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:8816-8820(1991).
 CC -!- FUNCTION: INVOLVED IN SPOROPHYTIC SELF-INCOMPATIBILITY SYSTEM (THE INABILITY OF FLOWERING PLANTS TO ACHIEVE SELF-FERTILIZATION), PROBABLY ACTING IN COMBINATION WITH S-LOCUS-SPECIFIC GLYCOPROTEINS. INTERACTION WITH A LIGAND IN THE EXTRACELLULAR DOMAIN TRIGGERS THE PROTEIN KINASE ACTIVITY OF THE CYTOPLASMIC DOMAIN.
 CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN = ADP + A PHOSPHOPROTEIN.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- TISSUE SPECIFICITY: PREDOMINANTLY IN THE PISTIL AND ANTHHER.
 CC -!- POLYMORPHISM: THERE ARE A NUMBER OF DIFFERENT S ALLELES IN B.OLERACEA, POSSIBLY PROVIDING THE RECOGNITION SPECIFICITY.
 CC -!- SIMILARITY: THE EXTRACELLULAR DOMAIN IS SIMILAR TO S-LOCUS GLYCOPROTEINS OF BRASSICA, WHILE THE INTRACELLULAR DOMAIN IS A SER/THR-PROTEIN KINASE RELATED TO RAF KINASES.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 DB EMBL; M76647; AAA33000.1; ALT_TERM.
 DR HSSP; P11362; 1FGI.
 DR InterPro; IPR001480; B_lectin.
 DR InterPro; IPR000719; Euk_kinase.

DR InterPro; IPR003609; Pan_app.
 DR InterPro; IPR002290; Ser_thr_kin_actsite.
 DR InterPro; IPR000858; Slocus_glycop.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00954; S_locus_glycop; 1.
 DR SMART; SM00108; B_lectin; 1.
 DR SMART; SM00473; PAN_AP; 1.
 DR SMART; SM00221; STYKC; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transferrase; Serine/threonine-protein kinase; Signal; ATP-binding; Transmembrane; Receptor; Glycoprotein; Self-incompatibility.
 FT SIGNAL 1 32
 FT CHAIN 33 849
 FT PUTATIVE SERINE/THREONINE KINASE RECEPTOR.
 FT DOMAIN 33 446
 FT TRANSMEM 447 466
 FT DOMAIN 467 849
 FT CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 528 779
 FT PROTEIN KINASE.
 FT NP_BIND 534 542
 FT ATP (BY SIMILARITY).
 FT BINDING 556 556
 FT ATP (BY SIMILARITY).
 FT ACT_SITE 653 653
 FT CARBOHYD 47 47
 FT CARBOHYD 120 120
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 196 196
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 260 260
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 314 314
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 389 389
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 442 442
 FT N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 849 AA; 97231 MW; 7E156059EDDF4370 CRC64;

Query Match 3.9%; Score 92.5; DB 1; Length 849;
 Best Local Similarity 21.6%; Pred. No. 13;
 Matches 77; Conservative 51; Mismatches 107; Indels 121; Gaps 24;

Qy 105 LVHGETPAVIRKTSIRLDETWTIKGPOVYIQTETDPIFLMG-----NEK---GMLTK 157
 Db 24 LIHPALSIYINT-----LSSTESLTISSNKTIV---SPGSIFEVGFRTNSRWYLGMYK 75
 Qy 158 K-DAKOLEYAAKQFTPLSLFSD-----LDNRNTPLWSSRPHNPMYVLPFIMHGK 205
 Db 76 KYSDRTYVWVANRDNPLSNAIGTLKISGNLVLDDHNSKPVWNTN-----LTRGN 125
 Qy 206 PNRSP-----NTPSHEAKQFT-----PNEFRAPELKQVSVKVAEADLW 245
 Db 126 -ERSPVVAELLANGFNVRDSSNNDASEYLWQSFYDPTDTLLPENKLYNLT----- 177
 Qy 246 GTDSDLWFGYTOQSHWQIFNGKNSRPRVRHYDQ-----PEIFLTQ---PVYSDLPWDGK 296
 Db 178 GLNREL-----TSWRSSDDPSSGNF---SYKLETQSLPEFYLSRENFPMHRSGPWNG- 226
 Qy 297 VMTGNGAVH-----HSNGESAKLSRWNRAY-----LMAGMEWKNLTVP--RI 339
 Db 227 IRFSGIPEDQKLSVMYNFENNEEVAYFTFTMNNFSRLTLISEGYFQRLTWPYSIRI 286
 Qy 340 WGRIFKEGSGGSDPNPDILDY-----YGYDVRFLYOLENKNKNSISGTVR--YNPRS 389
 Db 287 WNRFW-----SSPVD-POCDYIMCGPAYCDV-----NTSPVCNCIQGFNPRN 329

RESULT 12
 YL70_ARCFU STANDARD; PRT; 324 AA.
 AC O28112; 2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN AF2170.
 GN AF2170.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;


```

OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Krelavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -!- SIMILARITY: CONTAINS 2 KELCH REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000955; AAB89093.1; -
DR TIGR; AF2170; -
DR InterPro; IPR001798; Kelch.
DR Pfam; PF01344; Kelch; 2.
KW Hypothetical protein; Repeat; Complete proteome.
FT REPEAT 229 276
FT REPEAT 277 323
FT SEQUENCE 324 AA; 36025 MW; C445388CFEB96E45 CRC64;
SQ
Query Match 3.9%; Score 92; DB 1; Length 324;
Best Local Similarity 20.1%; Pred. No. 3.7;
Matches 54; Conservative 44; Mismatches 72; Indels 98; Gaps 18;
QY 224 EFRAPELKQVSVKVKAAKD-----LWGTDSDFQREAAVWDGGEILIFGCT-----VFENGKYSPTDQIL 112
DB 63 EHPNP-NLEI-----LLSSEDFQREAAVWDGGEILIFGCT-----VFENGKYSPTDQIL 112
QY 276 DYQPEIFLTPYISDLP-----W-DGKVRMI-----GMGAVHSHNGESAKLSRS 318
DB 113 SFNPKLERLVRVNLASLPHTSDVAAVWGDSDRVYIFLNNSECEVYAFYPSNESFAKLDSV 172
QY 319 -----W-NRAYL-----MAGMEKNTLVMPRIWGRIFKEGSGSQ 351
DB 173 CPTEHPGCVSHSVYWGKAYFCGEGVASFDPMGGFKW--IAFTDRVWVRAATVADG-- 228
QY 352 PDNDPDLIDYGYGDRVFLYQLENKSNISGT-----VRYNPRSGKGAOLDYVYPLGKG-- 405
DB 229 -----YFAFGSSGTAETKDEIRFPNPKTGE-LCEMRTKLPVARGQA 270
QY 406 --ISG-YFOIF-----QYQGSLIDYNH 425
DB 271 VAVGGEIYIFGGYTKDGYANEILRDYD 298
RESULT 13
R114.HUMAN
ID R114.HUMAN STANDARD; PRT; 1158 AA.
AC P48552;
DT 01-FEB-1996 (Rel. 33, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NUCLEAR FACTOR RIP140 (NUCLEAR RECEPTOR INTERACTING PROTEIN 1).

```

```

GN NR1P1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=95369246; PubMed=7641693;
RA Cavailles V., Dauvois S., L'Horsset F., Lopez G., Hoare S.,
RA Kushner P.J., Parker M.G.;
RT "Nuclear factor RIP140 modulates transcriptional activation by the
RT estrogen receptor."
RL EMBO J. 14:3741-3751(1995).
RN [2]
RC SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21."
RL Nature 405:311-319(2000).
CC -!- FUNCTION: MODULATES TRANSCRIPTIONAL ACTIVATION BY THE ESTROGEN
CC RECEPTOR.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X84373; CAA59108.1; -
DR EMBL; AL163207; CAB90396.1; -
DR MIM; 602490; -
KW Transcription regulation; Nuclear protein.
FT CONFLICT 124 124 P -> R (IN REF. 1).
FT CONFLICT 448 448 R -> G (IN REF. 1).
FT CONFLICT 721 726 NRGKSE -> TGRVK (IN REF. 1).
FT CONFLICT 1080 1080 T -> A (IN REF. 1).
FT SEQUENCE 1158 AA; 126941 MW; 81FC424968E9A5F6 CRC64;
SQ
Query Match 3.9%; Score 92; DB 1; Length 1158;
Best Local Similarity 19.3%; Pred. No. 21;
Matches 104; Conservative 70; Mismatches 150; Indels 214; Gaps 26;
QY 3 VSLSTLTLILSCFALLAQQAQVNPVAFVDEVESEND-----LQDNE 48
DB 458 VSLDNFTQSLNTW-----DPKVPDVIKEDQDTSKNSKLSHOKVTLQLLLGHKNE 510
QY 49 LPIDVQSATQ-----SASTDTANPLDEHEPELYTTA-----L 80
DB 511 ENVEKNTSPQGVHNDVSKENTQNYARTSVIESPSTNRTTPVS--TPPLTSSKAGSPINL 568
QY 81 ENKTMLIN-----CSALNQDIMRLACYDTL-----VHGETPAVIR-----TKRSI 120
DB 569 SQHSLVKNWSPPYVCSTQSEKLTNTASNHMDLTKSKDPPGKPAQNEGAQNSATFSAS 628
QY 121 RDETTWQTKGPKQVIYQETTDPIFLMGNEKGMCLKAKQLEYAAKQFTPLSLSLFDLD 180
DB 629 KLIQNLQAQCGMQSSMSVSEQRFSKQLLTGN-----TDK-----PIGM---ID 667

```

[illegible]

Qy 421 -----IDYNH 425
:|:|
Db 417 EKASREQOLDYSH 429

Search completed: November 30, 2001, 14:27:00
Job time: 547 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:17:45 ; Search time 36.79 Seconds
(without alignments)
270.358 Million cell updates/sec

Title: US-09-787-083-6
Perfect score: 2360
Sequence: 1 MKVSLSTLTLILSCFAIA.....YNHEATSGVGLMLNDWMGL 442

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100.5	4.3	682	3	US-08-481-435-6
2	100.5	4.3	1726	2	US-08-609-049A-30
3	100.5	4.3	1726	4	US-09-170-996-30
4	98.5	4.2	1658	2	US-08-609-049A-13
5	98.5	4.2	1658	4	US-09-170-996-13
6	90.5	3.8	1024	4	US-09-091-117-5
7	89.5	3.8	857	1	US-07-717-331F-2
8	88	3.7	537	2	US-08-633-879C-2
9	87.5	3.7	503	1	US-07-946-497-2
10	87.5	3.7	503	1	US-08-483-322-2
11	87.5	3.7	503	2	US-08-478-882-2
12	87.5	3.7	666	4	US-08-961-083-2
13	86	3.6	781	1	US-08-373-134D-2
14	86	3.6	781	2	US-09-114-637-2
15	86	3.6	1088	3	US-08-633-768A-1
16	85	3.6	816	1	US-07-731-157A-4
17	85	3.6	816	1	US-08-229-444B-2
18	85	3.6	816	2	US-08-541-780-4
19	85	3.6	985	5	PCT-US96-03916-6
20	85	3.6	985	5	PCT-US96-03916-66
21	84.5	3.6	774	3	US-08-902-632-2
22	84.5	3.6	774	3	US-09-073-354-1
23	84.5	3.6	774	3	US-08-656-005A-1
24	84.5	3.6	774	4	US-09-073-259-1
25	84.5	3.6	774	4	US-09-363-095-1
26	84.5	3.6	774	4	US-09-418-027-1
27	84	3.6	522	6	RE34606-6

Query Match 4.3%; Score 100.5; DB 3; Length 682;
Best Local Similarity 20.4%; Pred. No. 0.11; Indels 103; Gaps 24;
Matches 92; Conservative 68; Mismatches 189;

ALIGNMENTS

RESULT 1
US-08-481-435-6
; Sequence 6, Application US/08481435
; Patent No. 6027906
; GENERAL INFORMATION:
; APPLICANT: Balganes, Tanjore S
; APPLICANT: Town, Christine
; TITLE OF INVENTION: No. 6027906el Polypeptides
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,435
; FILING DATE: 10-JUL-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IN 580/MAS/94
; FILING DATE: 01-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9404072-2
; FILING DATE: 24-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Steiner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-151
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 682 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-481-435-6

Sequence 9, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 145, App
Sequence 48, Appli
Sequence 4, Appli
Sequence 7, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 4, Appli

US-08-609-049A-30

Query Match	4.3%	Score 100.5;	DB 2;	Length 1726;
Best Local Similarity	18.6%;	Pred. No. 0.54;		
Matches	98;	Conservative 79;	Mismatches 190;	Indels 161; Gaps
QY	1	MKVSLS-----TLTSLSCFAILAIQAAKAVNPVAFVD-----EVRSENDLQGD	46	
Db	465	VKSIIEIGQLPVTFCDVSSVTEIILIMQALSWHDLNQVDGVSYLLKVCQGEVQLN	524	
QY	47	NELPIDVQSATQSASTDTANPLDEHEPELYTTALENKTMLINGSALNQDLMRLACQD	106	
Db	525	NHCLGSHEHQNCRKWD-----EIKLQLLTLSAMCONLARTAEDD---	565	
QY	107	HGETPAVIKTKSRIRLDETTWQRIKGPQVYQETDPIF-----	146	
Db	566	--EAP-----VDLNKYLQTEKPYKEYMTRHPVEELDSYHYQVELALQTNQHR	615	
QY	147	-----LMGNEKGLTKKDAKQLEYAA-----KQFTPLSFLDORNNTPLWS	188	
Db	616	DOVIKAVRKICSALDGVETPSVEA-VKLLKRAVNLPRNKSAADVTSLSGSDTRK	674	
QY	189	SRHPNPMVILP-----IPWHGPNR-----SPWTPSHEARQFTPNFEAP	234	
Db	675	LNPNPQVQSMDLHTTRIYDLLRLHANSRSCGCPGRSNIKEAWTATE-----	726	
QY	235	SVKVKAAEDL---WGTDSDLWFGVTOQSHQIENGKN-SRPF---RVHDYQPEIF	285	
Db	727	QFTVYAAHGISSNVWSNYEKYLLCSLH---NGKDLFKPIQSKKVGTYKNFY	782	
QY	286	-----PV-YSDLPWDGKVRMIGCAVHHSNGESAKLSRSWN-----	324	
Db	783	ELIIFPIQISOLPLESVLHLLTFGLVNLQSSGSPDSNKQKGPALGVSLTLF	842	
QY	325	MAGMEWNKLTVMPRINGRIFKEGSGQPDNDPILDYGYGVDRFLYOLENKEN	384	
Db	843	TCG-----TKLLYLV-----TSSHNTSIIPAIPKKSVMERIVLQDFP	899	
QY	385	YNPRSGKALQLDVYVPLGKIGISGYFIQFYQGSGLIDYNHEATSGV	432	
Db	890	TSPQIDRNIIQOKLETFLSDIKG-----KLDDIHRDSSFL	927	

```

RESULT 3
US-09-170-996-30
; Sequence 30, Application US/09170996
; Patent No. 6291220
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Moliz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 6291220el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/170.996
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/609,049
; FILING DATE: 29-FEB-1996

```

ATTORNEY/AGENT INFORMATION:
NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 2307K-063700US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 1726 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-170-996-30

Query Match 4.3%; Score 100.5; DB 4; Length 1726;
Best Local Similarity 18.6%; Pred. No. 0.54;
Matches 98; Conservative 79; Mismatches 190; Indels 161; Gaps 21;
QY 1 MKVSL- ----TLTSLSCFAILAIQAKAVNPVAFVD-----EVSENDLGD 46
Db 465 VKVSIIEGLQLPVFTCDVSSTVEIIMQALSWHDDLNQVDVGSYILKVCQGEVLQN 524
QY 47 NELPIDVQSATSDTANPLDEHEPELYTTALENKTMLINCSALNQDIMRLACYDTLV 106
Db 525 NHCGLSHEHIQCRKWD-----EIKLQLLTLSAMCQNLARTAEED--- 565
QY 107 HGETPAVITKRSIRLDETITWQTIKGPQVYVQETTDPIF----- 146
Db 566 --EAP-----VDLNKLYIQIEKPYKEVMTRHPVEELDSYHYQVELALQTNQHRV 615
QY 147 -----LMGNEKGLMTKKDAKOLEVAA-----KQFTPLSLSFDDLDRNTPLS 188
Db 616 DQVIKAVKICSDALDGVETPSVTEA-VKKLRVNLPRNKSADVTSLSGSDTRKNTKGS 674
QY 189 SRPHNPWYLP-----IFMHGKPNR-----SPNTPSHEARQFTPNFEAPKLFQV 234
Db 675 LNPENPVQVSMDDLHTRIYDLRLHANSRCSGCPGRSNIKEAWTATE-----QL 726
QY 235 SVKVKAAEDL---WGTDSDLWFGYTOQSHWQIFNGKN-SRPF---RVHDYQPEIFLTQ-- 285
Db 727 QFTVYAAHGISSNWNYSKYLICSLSH-----NGKDLFKPIQSKKVGTYKNFFYLKWD 782
QY 286 ----PV-YSDLPWDGKVRMIGMGAHVHNSGESAKLSRSWN-----RAYL 324
Db 783 ELIIFPIQISQPLSVLHLLTFLGVNLQSSGSPDSNKQKRGPEALGVSLTLDFDKREL 842
QY 325 MAGMEKNLTVMPIRWGRIFKEGSGSQDDNDPILDYGYGDVRFYQLENKSNISGTVR 384
Db 843 TCG-----TKLLYLW-----TSSHTNSIPGAIPKKSVMYMERIVLQVDFPSPAFDIY 889
QY 385 YNPRCKGALOLDYVYPLCKGISGYFQIFQGYGQSILIDYNHEATSEFV 432
Db 890 TSPQIDRNIQDKLETLESIDK-----KLLDIHRDSSFGL 927

RESULT 4
US-09-609-049A-13
Sequence 13, Application US/08609049A
Patent No. 5948664
GENERAL INFORMATION:
APPLICANT: Williams, Lewis T.
APPLICANT: Molz, Lisa
APPLICANT: Chen, Yen-Wen
TITLE OF INVENTION: No. 5948664el PI 3-Kinase Polypeptides
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA

ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/609,049A
FILING DATE: 29-FEB-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 2307K-063700US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1658 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-609-049A-13

Query Match 4.2%; Score 98.5; DB 2; Length 1658;
Best Local Similarity 18.9%; Pred. No. 0.81;
Matches 101; Conservative 78; Mismatches 195; Indels 159; Gaps 23;
QY 5 LSTLTSLSCFAILAIQ---QAKAVNPVAFVDEYRS-----ENDLGDQNE 48
Db 381 LSPVTQVRNMGENASVKVSIIEGLQLPVFTCDVSSTVEIIMQALCWVHDDLNQ--- 437
QY 49 LPTDVOS-----ATQASASTDTANPLDEHE--PELYTTALENKTMLINCSALNQDIMRLAC 101
Db 438 --VDVGSYILKVCQGEVLQNHCGLSHEHIQCRKWDTEIKLQLLTLSAMCQNLARTAE 495
QY 102 YDPLVGHETPAVITKRSIRLDETITWQTIKGPQVYVQETTDPIF----- 146
Db 496 DD-----EAP-----VDLNKLYIQIEKPYKEVMTRHPVEELDSYHYQVELALQTN 542
QY 147 -----LMGNEKGLMTKKDAKOLEVAA-----KQFTPLSLSFDDLDRN 183
Db 543 QHRADVQVIKAVKICSDALDGVETPSVTEA-VKKLRVNLPRNKSADVTSLSGSDTRKN 601
QY 184 TPLWSSRPHPMTYLP-----IFMHGKPNR-----SPNTPSHEARQFTPNFEAPRAPE 229
Db 602 STKGSLNPNPVQVSMDDLHTTAIYDLRLHANSRCSGCPGRSNIKEAWTATE----- 656
QY 230 LKFOVSVKVKAAEDL---WGTDSDLWFGYTOQSHWQIFNGKN-SRPF---RVHDYQPEIF 282
Db 657 ---OLOFTVYAAHGISSNWNYSKYLICSLSH-----NGKDLFKPIQSKKVGTYKNFFY 709
QY 283 LTQ-----PV-YSDLPWDGKVRMIGMGAHVHNSGESAKLSRSWN----- 320
Db 710 LIKWDELIIFFIPIQISQPLSVLHLLTFLGVNLQSSGSPDSNKQKRGPEALGVSLTLFD 769
QY 321 -RAYLMAGMEKNLTVMPIRWGRIFKEGSGSQDDNDPILDYGYGDVRFYQLENKSN 379
Db 770 FKRELTGCG-----TKLLYLW-----TSSHTNSIPGAIPKKSVMYMERIVLQVDFPSPA 816
QY 380 SGTVRYNPRCKGALOLDYVYPLCKGISGYFQIFQGYGQSILIDYNHEATSEFV 432
Db 817 FDIYTSPOIDRNIQDKLETLESIDK-----KLLDIHRDSSFGL 859

RESULT 5
US-09-170-996-13
Sequence 13, Application US/09170996
Patent No. 6291220
GENERAL INFORMATION:

APPLICANT: Williams, Lewis T.
APPLICANT: Molz, Lisa
APPLICANT: Chen, Yen-Wen
TITLE OF INVENTION: No. 6291220el PI 3-Kinase Polypeptides
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/170,996
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/609,049
FILING DATE: 29-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 2307K-0637000S
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1658 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-170-996-13

Query Match 4.2%; Score 98.5; DB 4; Length 1658;
Best Local Similarity 18.9%; Pred. No. 0.81;
Matches 101; Conservative 78; Mismatches 195; Indels 159; Gaps 23;

Qy 5 LSTLTLSLSCFAILAIQ---QAKAVPNPVPFVDEVR-----ENDLGQDNE 48
Db LSPVTQVRNMGENASVKVSIIEGLQLPVTFTCDVSTVEIIMQALCWVHDDLQ--- 437
Qy 49 LPIDVQS-----ATQASTDTANPLDEHE--PELYTTALENKTMLINGSALNQDIMRLAC 101
Db --VDVGSYILKVGQEEVLQNNHCLSGSHEHQNCRKWDTEIKLQLLTLSAMCQNLARTAE 495
Qy 102 YDTLVHGETPAVTKRSIRDETQWIKGPKQPVVYQETTDPIF----- 146
Db -----EAP-----VDLNKYLQKEKPYKEMIRHPVEELDSYHYQVELALQFEN 542
Qy 147 -----LMGNEKMLTKDAKOLEYAA-----KQFTPLSLSFDLDRNN 183
Db QHRAVDQVIKAVRKICSLDGVETPSVTEA-VKLLKRAVNLPRNKSADVTSLSGSDTRKN 601
Qy 184 TPLWSSRRPHNPMYVLP-----IPMHGKPNR-----SPNTPSHEARQFTPNNEFRAPE 229
Db 602 STKGLSNPENVOVSMDLHTTAIVDLLRLHANSSRCSTGCPGRSRRNKEAWTATE----- 656
Qy 230 LKFOVSVKKAEDL---WGTDSDLWFGYTOQSHWQIFNGKN-SRPF---RVHDIQPIEF 282
Db 657 ---QLQFTVYAAHGSSNMVSNYKYYLICSLSH-----NGKDLFKPIOSKKGVTYKNEFY 709
Qy 283 LTQ-----PV-YSDLPDWDGKVRMIGMGAHVHSHNGESAKLSRSWN----- 320
Db 710 LIKWDELIIFIQISQLPESVILHTLFGVLNQSSGSSPDNSKQKRGPEALGKVSILTIFD 769

Qy 321 -RAYLMAGMEWKNLTVMPIRNGRIFKEGSGSQDPNDPILDYGYGVDYRFLYQLENKSN1 379
Db 770 PKRFLTCG-----TKLLYLW-----TSSHTNIPGAIPKKSVMERIVLQVDFPSPA 816
Qy 380 SGTVYVNPGRSGKALQLDYVYPLGKIGSGYQIFOGYQSLIDYNHEATSEGV 432
Db 817 FDIYVTSPOIDRNIIQDKLETLESIDKG-----KLLDIHRDSSFG 859
RESULT 6
US-09-091-117-5
Sequence 5, Application US/09091117
Patent No. 6171589
GENERAL INFORMATION:
APPLICANT: The University of Melbourne
TITLE OF INVENTION: Mycoplasma Recombinant Polypeptides and
TITLE OF INVENTION: Vaccines
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: GREENLEE, WINNER and SULLIVAN P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: United States of America
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/091,117
FILING DATE: 12 JUNE 1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU96/00803
FILING DATE: 13-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PN7127
FILING DATE: 13-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: WINNER, Ellen P.
TELECOMMUNICATION INFORMATION:
TELEPHONE: +1 303 499 8080
TELEFAX: +1 303 499 8089
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1024 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Mycoplasma genitalium
US-09-091-117-5

Query Match 3.8%; Score 90.5; DB 4; Length 1024;
Best Local Similarity 20.2%; Pred. No. 2.6;
Matches 86; Conservative 61; Mismatches 172; Indels 107; Gaps 22;

Qy 1 MKVS-LSTLTLSLSCFAILAIQAKAVPNPVPFVDEVRSENDLGQDNELP1-DVQSATQ 58
Db 1 MKLSTITTTICLSISGAFCTTAI---ALPTTVALLKNHQOQNTKEQQN--PIKDIRFGLN 54
Qy 59 SASDTPANPLDEHEPELYTTALENKTMLINGSALNQDIMRLACYDTLVHGETPAVTKR 118
Db 55 NVQVPNTIPL--HQTVVEVT--NNKAIYDYKDAPOKFFL-----AKSALNNKL 98
Qy 119 SIRLDETQWIKGPKQPVVYQETTDPIFLMGNEKMLTKDAKOLEYAAKO-FTPLSLSF 177
Db 99 QVEFDKFLRT-----GVINALNADLKEWIDQTLFIPNQSF 135
Qy 178 DLDRNNTPLWSSRRPHNPMYVLP-IFMHGKPNRSP-NTPSHEARQFTPNNEFRAPELKFV 235

Db 136 DLSANKNLTLNQSSEVSLDLEIFITNFSKNOPLKLPEDGSVVVANE-----SYTYSVK 191
QY 236 VKYAAEDLWGTDSDLWFGYTOQSHWOIFNGKNSRPRVRHDOPEI-FLTQPVYSDI-PW 293
Db 192 ATLQKLVITYSRADHSVGSYIAITVSLNGKTQDNFSENPFSKSNINFAFKVYNALNPF 251
QY 294 DGKVRMIGCAVHHS--NGESAK-----LSRSWNRAYLMAGMEWKNLTVMPRIWGRIFK 345
Db 252 EAOGYLVGCKFKLVQKVNADVDKNDINNHIETOFNVAKITA-----TLGKAEK 300
QY 346 ---EGSGSOP-----DNDPILDVY--GYGDV-----REFLYOL-- 373
Db 301 QFGEHNGOPLSLKLVSLGNLEFKQLFNVRPGLGDFVSDLIQSSQSSNKKTKVQLLF 360
QY 374 ENKSN1 379
Db 361 ENKTTI 366

RESULT 7

US-07-717-331F-2
; Sequence 2, Application US/07717331F
; Patent No. 5484905
; GENERAL INFORMATION:
; APPLICANT: June Nasrallah; Michael Nasrallah; and Joshua
; APPLICANT: Stein
; TITLE OF INVENTION: A Receptor Protein Kinase Gene
; TITLE OF INVENTION: Encoded At The Self-Incompatibility Locus
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yahwak & Associates
; STREET: 25 Skytop Drive
; CITY: Trumbull
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06611
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07717,331F
; FILING DATE: June 19th 1991
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: George M. Yahwak
; REGISTRATION NUMBER: 26,824
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203)268-1951
; TELEFAX: (203)268-1951
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 857 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-717-331F-2

Query Match 3.8%; Score 89.5; DB 1; Length 857;
Best Local Similarity 21.4%; Pred. No. 2.4;
Matches 76; Conservative 50; Mismatches 110; Indels 119; Gaps 23;

QY 105 LVHGETPAVTKRSRLDETHTWTKGPOVYQYTTDPFLMG-----NEK---GMLTK 157
Db 24 LTHPALSIYINT-----LSSTESLATISSNKTLY---SPGSIIEVGFRTNSRWYLGMYK 75
QY 158 K-DAKQLEYAAKQFTPLSLSD-----LDNRNTPLWSSRPHNPVYLPPIEMHCK 205
Db 76 KYSDRYVWVANDNPLNSAIGTLKISGNLVLLDHSNKPVMWTN-----LTRGN 125
QY 206 PNRSP-----NTPSHEARQFT-----PNEFRAPELKFQVSVKVAADLW 245

Db 126 -ERSPVVAELANGNFVMRDSSNNDASEYLWQSFDPYPTDTLLPEMKLYNLKT----- 177
QY 246 GTDSDLWFGYTOQSHWOIFNGKNSRPRVRHDO-----PEIFLTQ---PVYSDLPWDGK 296
Db 178 GLNRFL-----TSNRSSDDPSGNF---SYKLETSQSLPEFYLSRENFMHRSRGPWNG- 226
QY 297 VRMIGMAVH-----HNGESAKLSRSWNRAY-----LMAGMEWKNLTVMP--RI 339
Db 227 IRESGIPEDQKLSYMYVNFETENNEEVAYTFRTMTNNSFSRLTLISEGYFORLTWYPSIRI 286
QY 340 WGRIFEGSGSQPDD-----NPDILDYGYGDVRFYQLENKNSISGTVR--YNPRS 389
Db 287 WNRFSWSPVDRQCDTVIMCGP-----YAYCDV-----NTSPVCNIOGFNPRN 329

RESULT 8

US-08-633-879C-2
; Sequence 2, Application US/08633879C
; Patent No. 5928922
; GENERAL INFORMATION:
; APPLICANT: Kivirikko, Kari I.
; APPLICANT: Pihlajaniemi, Taina
; APPLICANT: Helaakoski, Tarja I.
; APPLICANT: Annunen, Pia P.
; APPLICANT: Nissi, Ritva K.
; APPLICANT: No. 5928922elainen, Minna K.
; TITLE OF INVENTION: '2 SUBUNIT OF PROLYL-4-HYDROXYLASE
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING SUCH SUBUNIT AND
; TITLE OF INVENTION: METHODS FOR PRODUCING THE SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,879C
; FILING DATE: 10-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B.
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0041-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 537 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-633-879C-2

Query Match 3.7%; Score 88; DB 2; Length 537;
Best Local Similarity 21.4%; Pred. No. 1.6;
Matches 100; Conservative 52; Mismatches 145; Indels 170; Gaps 26;

TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-483-322-2

Query Match 3.7%; Score 87.5; DB 1; Length 503;
Best Local Similarity 18.4%; Pred. No. 1.6;
Matches 57; Conservative 31; Mismatches 97; Indels 125; Gaps 14;

QY 40 ENDLQDNELPIDVQSATQSASTDTANPLDEHEPELYTTALENKTMILNCSALNQDIMRL 99
Db 169 QEDIDASNIIDEVSS-----GSTIEKSTPEGYILHTDLP---SQTGDRD--- 212
QY 100 ACYDTLVHGETPAVTKRSI-----RLDETIVQIKGPKQVYQETTDPIFLMGNEKG 153
Db 213 ---DAFFIGSTLATIATTPWVSAHTKQNCQERTQWNPVLSNPEVLLQTT----- 258
QY 154 MLTKKDAKQLEYAAKQFTPLSLSFDRNTPL-----WSSRPHNPMVLPVIFMHGKPNRS 209
Db 259 -----RMT-----DIDRNSTSAHGENWTQEPQPPF----- 283
QY 210 PNTPSHEARQFTNEFRAPELKFQSVKVAEEDLW-----GTDSDLWFGVYTOQS 259
Db 284 -NNHEYQDEEETH-----ATSTWADPNSTTEEAATQEKWF-----EN 322
QY 260 HMQIFNGKN-----SRPFRVHDYQPEIFLTQPVYSDLPW-----DGKVRMIGMG 303
Db 323 EMQ---GKNPPTPSDSHVTEGTTASAHNNHPSQRTTQSQEDVSWTDFDPIISHPMGQG 379
QY 304 AVHHSNGESA 313
Db 380 HOTESKGHSS 389

RESULT 11
US-08-478-882-2
Sequence 2, Application US/08478882
Patent No. 5885575
GENERAL INFORMATION:
APPLICANT: HERRLICH, Peter
APPLICANT: PONTA, Helmut
APPLICANT: GUENTHER, Ursula
APPLICANT: MATZKU, Siegfried
APPLICANT: WENZL, Achim
TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA
TITLE OF INVENTION: SEQUENCES CODING THESE ANTIBODIES AGAINST THESE PROTEINS,
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478.882
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/946.497
FILING DATE: 19921109
ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16915/145
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-882-2

Query Match 3.7%; Score 87.5; DB 2; Length 503;
Best Local Similarity 18.4%; Pred. No. 1.6;
Matches 57; Conservative 31; Mismatches 97; Indels 125; Gaps 14;

QY 40 ENDLQDNELPIDVQSATQSASTDTANPLDEHEPELYTTALENKTMILNCSALNQDIMRL 99
Db 169 QEDIDASNIIDEVSS-----GSTIEKSTPEGYILHTDLP---SQTGDRD--- 212
QY 100 ACYDTLVHGETPAVTKRSI-----RLDETIVQIKGPKQVYQETTDPIFLMGNEKG 153
Db 213 ---DAFFIGSTLATIATTPWVSAHTKQNCQERTQWNPVLSNPEVLLQTT----- 258
QY 154 MLTKKDAKQLEYAAKQFTPLSLSFDRNTPL-----WSSRPHNPMVLPVIFMHGKPNRS 209
Db 259 -----RMT-----DIDRNSTSAHGENWTQEPQPPF----- 283
QY 210 PNTPSHEARQFTNEFRAPELKFQSVKVAEEDLW-----GTDSDLWFGVYTOQS 259
Db 284 -NNHEYQDEEETH-----ATSTWADPNSTTEEAATQEKWF-----EN 322
QY 260 HMQIFNGKN-----SRPFRVHDYQPEIFLTQPVYSDLPW-----DGKVRMIGMG 303
Db 323 EMQ---GKNPPTPSDSHVTEGTTASAHNNHPSQRTTQSQEDVSWTDFDPIISHPMGQG 379
QY 304 AVHHSNGESA 313
Db 380 HOTESKGHSS 389

RESULT 12
US-08-961-083-2
Sequence 2, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 666 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-2

Query Match 3.7%; Score 87.5; DB 4; Length 666;
Best Local Similarity 20.1%; Pred. No. 2.6; Mismatches 188; Indels 97; Gaps 24;
Matches 88; Conservative 64; Mismatches 188; Indels 97; Gaps 24;
QY 25 KAVPNPFAFVDEVRSENDL-GQDNELPIDVQSATQS-----ASTDTANPLDEHEPE 74
DB 1 KIYDNKNLIADLGSERRVNAQANDIPTDLVKAIVSIEDHREFDHRGIDTIRILGAFLRN 60
QY 75 LYTTALENKTMLNCSALNQDIMRLACVDTLVHGGETPAVIKTSIRLDETIWQTIKGP 134
DB 61 LOSNSLOGG-----STLTQOLIKLYFSTSTSDQIS-----RKAQEAWLAIQLEQ 106
QY 135 OVVOYQET-----TDPIFLMGNEKGLTK-----DKAKOLE-----YAAKQFTPL 173
DB 107 KATQKEILTYINKYMGNGVNGWTAQNYGKDLNLSLPQLALLAGMPQAPNQIDPY 166
QY 174 S-LSPDLRNNTPWSSRPHNPVLPPIEMHGKPNRSPNTPSHEARQFTPNFEFRAPE--- 229
DB 167 SHPEAAQDRRLVL--SEMKNGYI---SARQYEKAVNTPITDGLQSLKSASNPAYMD 220
QY 230 --LKFQVS-VKVAEEDLWGTDSLWFGYTO--QSH-WQIFNGKNSRPFVHDYQPEIFL 283
DB 221 NYLKEVINQVEETGYNLLTTGMVYTNVDQAQKHLWDIYNTDEYVAYPDDELQ----- 275
QY 284 TQPVYSDLPWCKVRMIGMGAHVHNS-----GESAKLSRSNNRAYLMAGMEWKNLT-VM 336
DB 276 VASTIVDVS-NGKV-IAQLGARHQSSNVSGINQAVENRQW-----GSTMKEPITDYA 326
QY 337 PRI-----WGRIFKEGSGQPDNDILDY-YGY-GDVRFYQLENKSNISGTVRYN- 386
DB 327 PALEYGVYDSTATIVHDEPNYPNGTNPVYNWDRGYFGNITLQYALQOSRNVPAVETLNK 386
QY 387 -----PRSGKGAQLQDY 398
DB 387 VGLNRAKTFLLGLGIDY 403

RESULT 13
US-08-373-134D-2
; Sequence 2, Application US/08373134D
; Patent No. 5780296
; GENERAL INFORMATION:
; APPLICANT: Kmiec, Eric
; APPLICANT: Holloman, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS TO PROMOTE
; TITLE OF INVENTION: HOMOLOGOUS RECOMBINATION IN EUKARYOTIC CELLS AND ORGANISMS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; APPLICATION NUMBER: US/09/114,637
; FILING DATE:

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/373,134D
FILING DATE: January 17, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Friebe, Thomas E.
REGISTRATION NUMBER: 29,258
REFERENCE/DOCKET NUMBER: 7991-007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 781 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-373-134D-2

Query Match 3.6%; Score 86; DB 1; Length 781;
Best Local Similarity 20.7%; Pred. No. 4.9; Mismatches 122; Indels 32; Gaps 9;
Matches 50; Conservative 37; Mismatches 122; Indels 32; Gaps 9;
QY 39 SENDLGOONELPIDVQS-ATQASDTDTANPLDEHEPELYTTALENKTMLNCSALNQDIM 97
DB 533 ARNDGTSQAPTSAFSGTGSALPDQPLAMDVASQTAFSTGLASIAPTAEAVGAREL 592
QY 98 RLACYDTLVHGGETPAVIKTSIRLDETIWQTI-----KGKPVVYQE-----TTDPI 145
DB 593 DSAC-----ASNDVPLRTLEARTAQLGOT-WSNLINRVFLSKTRARICMRDDQAPACEPV 647
QY 146 FLMGNEKMLTKKDAKOLEYAAKQFTPLSLF-DLDRNNTPLASSR-----PHNPMVLP 200
DB 648 RONTNQRTASKSLMNTVRKAAVINPFEGATMLDVGVDKLSALRQLRVTITPRKAVHVL-- 705
QY 201 FMHGKPNRSPNTPSHEARQFTPNFEFRAPELKFQSVKVAEEDLWGTDSLWFGYTOQSH 260
DB 706 -----NAYPSTVMH-AMHATADSTPAPESSQOQRAAERHPAEQEDADQDLFGEALQEH 758
QY 261 W 261
DB 759 W 759

RESULT 14
US-09-114-637-2
; Sequence 2, Application US/09114637
; Patent No. 5945339
; GENERAL INFORMATION:
; APPLICANT: Kmiec, Eric
; APPLICANT: Holloman, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS TO PROMOTE
; TITLE OF INVENTION: HOMOLOGOUS RECOMBINATION IN EUKARYOTIC CELLS AND ORGANISMS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; APPLICATION NUMBER: US/09/114,637
; FILING DATE:

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:16:59 ; Search time 72.04 Seconds
(without alignments)
454.475 Million cell updates/sec

Title: US-09-787-083-6
Perfect score: 2360
Sequence: 1 MKVSLSTLTLSLSCFAILA.....YNHEATSGVGLMLNDWMGL 442

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	2360	100.0	442	21	AAV85270 BASB034 amino acid
2	2352	99.7	442	21	AAV85268 BASB034 amino acid
3	2348	99.5	442	21	AAV85271 BASB034 amino acid
4	2330	98.7	442	21	AAV85269 BASB034 amino acid
5	822	34.8	370	21	AAV75156 Neisseria gonorrhoe
6	815	34.5	370	21	AAV75157 Neisseria meningit
7	815	34.5	370	21	AAV75158 Neisseria meningit
8	815	34.5	374	21	AAV70629 Neisseria meningit
9	812	34.4	375	21	AAV70628 Neisseria meningit
10	246.5	10.4	355	19	AAW98871 H. pylori GHPO 172
11	239.5	10.1	356	19	AAV10960 H. pylori ORF 07ap

12	156.5	6.6	253	18	AAW20760 H. pylori outer me
13	114	4.8	1686	19	AAW70991 Human class II p13
14	109.5	4.6	800	8	AAV70420 Sequence encoded b
15	109.5	4.6	822	13	AAV26021 Alkaline cellulase
16	108	4.6	824	21	AAV23180 Bacillus sp. KSM-S
17	107	4.5	157	18	AAW20538 H. pylori outer me
18	104.5	4.4	537	22	AAV91000 C glutamicum prote
19	104	4.4	798	21	AAV40925 Human ORF689
20	104	4.4	798	21	AAV90225 Human PGC-1 protel
21	104	4.4	798	22	AAV83944 Peroxisome prolif
22	100.5	4.3	682	17	AAV04359 S. pneumoniae peni
23	99.5	4.2	1726	18	AAV38756 Phosphatidyl inos
24	98.5	4.2	467	20	AAV13378 Amino acid sequenc
25	98.5	4.2	467	21	AAV94860 Human protein clon
26	98.5	4.2	467	22	AAV88401 Human membrane or
27	98.5	4.2	467	22	AAV80246 Human PRO260 prote
28	95.5	4.0	502	22	AAV90031 C glutamicum prote
29	95.5	4.0	516	22	AAV78876 C. glutamicum SRT
30	95	4.0	522	22	AAV95616 Human protein sequ
31	94	4.0	372	16	AAV69607 Gonococcal porin-5
32	94	4.0	761	20	AAV99084 Non-B, non-C, non-
33	94	4.0	1115	21	AAV93947 Amino acid sequenc
34	93.5	4.0	467	19	AAV75057 Human secreted pro
35	92.5	3.9	1717	22	AAV20498 Ostria nubialis
36	92	3.9	888	22	AAV70751 S cerevisiae apopt
37	91.5	3.9	857	13	AAV29814 S receptor kinase
38	91	3.9	1227	22	AAV81501 S. epidermidis ope
39	90.5	3.8	1024	18	AAV19604 Mycoplasma genital
40	90	3.8	460	21	AAV25456 Arabidopsis thalia
41	90	3.8	481	21	AAV25455 Arabidopsis thalia
42	90	3.8	522	22	AAV95513 Human protein sequ
43	90	3.8	564	21	AAV10457 T. matsutake pyran
44	90	3.8	564	21	AAV81952 Trichoderma deri
45	90	3.8	600	22	AAV75092 Human colon cancer

ALIGNMENTS

RESULT 1
AAV85270
ID AAV85270 standard; Protein; 442 AA.
XX
AC AAV85270;
XX
DT 29-JUN-2000 (first entry)
XX
DE BASB034 amino acid sequence #3.
XX
KW Moraxella catarrhalis infection; BASB034; diagnosis; staging;
KW vaccine; bacteriostatic; treatment; prevention; otitis media; pneumonia;
KW sinusitis; nosocomial infection; invasive disease; chronic otitis media;
KW hearing loss; antibacterial drug.
XX
OS Moraxella catarrhalis.
XX
PN WO200015802-A1.
XX
PD 23-MAR-2000.
XX
PF 14-SEP-1999; 99WO-EP06781.
XX
PR 14-SEP-1998; 98GB-0020002.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Ruelle J;
XX
DR WPI: 2000-271440/23.
XX
N-PSDB; AAA10702.
XX
PT Novel BASB034 polynucleotides and polypeptides from Moraxella
catarrhalis used to prepare vaccines against bacterial infections

XX PS Claim 3; Page 68; 106pp; English.

XX CC This sequence represents a Moraxella catarrhalis BASB034 polypeptide from

XX CC strain Mc2913. The invention relates to BASB034 polypeptides from

XX CC M. catarrhalis strains Mc2931, Mc2908, Mc2913 and Mc2969. The BASB034

XX CC polynucleotides and polypeptides may be employed as research reagents and

XX CC material for the discovery of treatments and diagnostics for diseases,

XX CC particularly human diseases. They are particularly used to diagnose and

XX CC treat M. catarrhalis infections. They can be used for diagnosis of

XX CC disease, staging of disease, or determining response of an infectious

XX CC organism to drugs. The polynucleotides may be used as a source for

XX CC hybridization probes, and for screening of genetic mutations, serotype,

XX CC organism or strain identification, identification of mutations in BASB034

XX CC sequences, and as components of arrays which are useful for diagnostic

XX CC and prognostic purposes. The polypeptides can be used to produce

XX CC antibodies. The polypeptides can also be used in vaccine formulations,

XX CC and to identify agonists and antagonists. The polypeptides, antibodies,

XX CC agonists and antagonists (which are bacteriostatic) are used for the

XX CC treatment and prevention of diseases such as otitis media in infants and

XX CC children, pneumonia in elderlies, sinusitis, nosocomial infections and

XX CC invasive diseases, and chronic otitis media with hearing loss. The

XX CC polypeptides, agonists and antagonists are also used for screening of

XX CC antibacterial drugs. The BASB034 products of the invention can be used

XX CC screen for new antibacterial compounds that may target resistant

XX CC bacteria.

XX SQ Sequence 442 AA;

Query Match 100.0%; Score 2360; DB 21; Length 442;

Best Local Similarity 100.0%; Pred. No. 1.3e-218;

Matches 442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKVSLSTLTSLSCFAILAIQAKAVPNPVADEVSRSENDLGDQNELPIDVQATQSA 60

DB 1 mkvslstltsilscfaiilaqakavpnvpafdevrsendlgdnelpidvqatqsa 60

QY 61 STDANPLDEHEPELYTALENKTMILNCSALNQDMLRACVDTLVHGETPAVITKRSI 120

DB 61 stdtanpldehepeilytalenktmlncsalnqdimrlacydtlvhgetpaviktksi 120

QY 121 RUDETIWOTIKGKPOVVOETDPIFLMGNEKGMTKKDAKOLEYAAKQFTPLSFSFDLD 180

DB 121 rldetiwtotikgkpoovyvoettdpiflmgnekgmtkkdakoleyaakqftplsfsfld 180

QY 181 RNTPTLWSSRPHNPVLPFLFMHGKPNRSPNTPSHEARQFTNEFRAPLKFQVSVKYKA 240

DB 181 rntptlwsrphnpmyvlpfmngkpnrsptnpshearqftnefrapelkfqsvkyka 240

QY 241 ABDLWGTOSDLWFGYTQSHWQIFNGKNSRPFVRHDIYQPEIFLTPQVYSDLPWDGKVRMI 300

DB 241 aedlwgtosdlwfgytqshwqifngknsrpfvrhdyqpeifltqpvysdlpwdgkvrmi 300

QY 301 GMGAVHHNGESAKLSRSWNRYALMAGMEWKNLTVMPTRIWGRIFKEGSGSQDDNPDIID 360

DB 301 gmgaavhshngesaklsrswrnraylmagmewknlvtmptriwgrifkegsgsqddnpdiid 360

QY 361 YGYGDVRFVLYOLENKSNTSGVRYNPRSGKALQDLYVPLGKISGYFQIFQGYGOSL 420

DB 361 ygygdvrfvlyolenksntsgvrynprrsgkalqldlyvypigklsyfyfqqgyggsi 420

QY 421 IDYNHEATSFYGLMLNDWMGL 442

DB 421 idynheatsfvglmndwmg 442

RESULT 2

AY85268

XX AY85268 standard; Protein: 442 AA.

AC AY85268;

XX

DT 29-JUN-2000 (first entry)

XX BASB034 amino acid sequence #1.

XX Moraxella catarrhalis infection; BASB034; diagnosis; staging;

XX vaccine; bacteriostatic; treatment; prevention; otitis media; pneumonia;

XX sinusitis; nosocomial infection; invasive disease; chronic otitis media;

XX hearing loss; antibacterial drug.

XX Moraxella catarrhalis.

XX WO200015802-A1.

XX 23-MAR-2000.

XX 14-SEP-1999; 99WO-EP06781.

XX 14-SEP-1998; 98GB-0020002.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Ruelle J;

XX WPI; 2000-271440/23.

XX N-PSDB; AAA10700.

XX Novel BASB034 polynucleotides and polypeptides from Moraxella

XX catarrhalis used to prepare vaccines against bacterial infections

XX Claim 3; Fig 2; 106pp; English.

XX This sequence represents a Moraxella catarrhalis BASB034 polypeptide from

XX strain Mc2931 (ATCC 43617). The invention relates to BASB034 polypeptides

XX from M. catarrhalis strains Mc2931, Mc2908, Mc2913 and Mc2969. The

XX BASB034 polynucleotides and polypeptides may be employed as research

XX reagents and material for the discovery of treatments and diagnostics for

XX diseases, particularly human diseases. They are particularly used to

XX diagnose and treat M. catarrhalis infections. They can be used for

XX diagnosis of disease, staging of disease, or determining response of an

XX infectious organism to drugs. The polynucleotides may be used as a source

XX for hybridization probes, and for screening of genetic mutations,

XX serotype, organism or strain identification, identification of mutations in

XX BASB034 sequences, and as components of arrays which are useful for

XX diagnostic and prognostic purposes. The polypeptides can be used to

XX produce antibodies. The polypeptides can also be used in vaccine

XX formulations, and to identify agonists and antagonists. The polypeptides,

XX antibodies, agonists and antagonists (which are bacteriostatic) are used

XX for the treatment and prevention of diseases such as otitis media in

XX infants and children, pneumonia in elderlies, sinusitis, nosocomial

XX infections and invasive diseases, and chronic otitis media with hearing

XX loss. The polypeptides, agonists and antagonists are also used for

XX screening of antibacterial drugs. The BASB034 products of the invention

XX can be used screen for new antibacterial compounds that may target

XX resistant bacteria.

XX SQ Sequence 442 AA;

Query Match 99.7%; Score 2352; DB 21; Length 442;

Best Local Similarity 99.3%; Pred. No. 7.6e-218;

Matches 439; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKVSLSTLTSLSCFAILAIQAKAVPNPVADEVSRSENDLGDQNELPIDVQATQSA 60

DB 1 mkvslstltsilscfaiilaqakavpnvpafdevrsendlgdnelpidvqatqsa 60

QY 61 STDANPLDEHEPELYTALENKTMILNCSALNQDMLRACVDTLVHGETPAVITKRSI 120

DB 61 stdtanpldehepeilytalenktmlncsalnqdimrlacydtlvhgetpaviktksi 120

QY 121 RUDETIWOTIKGKPOVVOETDPIFLMGNEKGMTKKDAKOLEYAAKQFTPLSFSFDLD 180

DB 121 rldetiwtotikgkpoovyvoettdpiflmgnekgmtkkdakoleyaakqftplsfsfld 180

QY 181 RNNTPWSSRPHNPMYVLPFIMHGKPNRSPNTPSHEARQFTNPEFRAPELKFOVSVKVA 240
 CC |||||||
 Db 181 rnnptlwssrphnmpmyvlpfimhggkpnrsnptpsheakqftnpefrapelkfgsvkvka 240
 CC |||||||
 QY 241 AEDLWGTDSLWFGYTOQSHWQIFNGKNSRPFVRHVDYQPEIFLTQPVYSDLPWDGKVRMI 300
 CC |||||||
 Db 241 aedlwgtdslwfgytqgshwqifngknsrpfvrhdyqpeifltqpvysdlpwdgkvrmi 300
 CC |||||||
 QY 301 GMGAVHHSNGESAKLSRSNRRAYLMAGMEWKNLTVMPRIGRIFKEGSGSQPDNDPILD 360
 CC |||||||
 Db 301 gmgavhhsngesaklsrsnraylmagmewknlvtmprigrifkegsgsqpdndpild 360
 CC |||||||
 QY 361 YGYGQVDFRFLYQLENKSNISGTVRYNPRSGKQALQLDYVYPLGKISGYFQIFGQCSL 420
 CC |||||||
 Db 361 yygygvdfrflyqlenksnsgtvrynprrsgkqalqldyvypylgkigsgyfqiifggyqgs 420
 CC |||||||
 QY 421 IDYNHEATSEFGVGLMLNDWMGL 442
 CC |||||||
 Db 421 idynheatsfvgvglmndwmgl 442
 CC |||||||
 RESULT 3
 AAY85271
 ID AAY85271 standard; Protein; 442 AA.
 AC AAY85271;
 DT 29-JUN-2000 (first entry)
 XX BAS034 amino acid sequence #4.
 DE Moraxella catarrhalis.
 KW vaccine; catarrhalis infection; BAS034; diagnosis; staging;
 KW sinusitis; nosocomial infection; treatment; prevention; otitis media; pneumonia;
 KW hearing loss; antibiotic drug.
 XX
 OS Moraxella catarrhalis.
 XX
 PN WO200015802-A1.
 XX
 PD 23-MAR-2000.
 XX
 PF 14-SEP-1999; 99WO-EP06781.
 XX
 PR 14-SEP-1998; 98GB-0020002.
 XX
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PI Ruelle J;
 XX
 DR WPI: 2000-271440/23.
 DR N-PSDB; AAL10703.
 XX
 PT Novel BAS034 polynucleotides and polypeptides from Moraxella
 CC catarrhalis used to prepare vaccines against bacterial infections
 PT
 PS Claim 3; Page 69; 106pp; English.
 XX
 CC This sequence represents a Moraxella catarrhalis BAS034 polypeptide from
 CC strain Mc2969. The invention relates to BAS034 polypeptides from
 CC M. catarrhalis strains Mc2931, Mc2908, Mc2913 and Mc2969. The BAS034
 CC polynucleotides and polypeptides may be employed as research reagents and
 CC material for the discovery of treatments and diagnostics for diseases,
 CC particularly human diseases. They are particularly used to diagnose and
 CC treat M. catarrhalis infections. They can be used for diagnosis of
 CC disease, staging of disease, or determining response of an infectious
 CC organism to drugs. The polynucleotides may be used as a source for
 CC hybridization probes, and for screening of genetic mutations, serotype,
 CC organism or strain identification, identification of mutations in BAS034
 CC sequences, and as components of arrays which are useful for diagnostic
 CC and prognostic purposes. The polypeptides can be used to produce
 CC antibodies. The polypeptides can also be used in vaccine formulations,

CC and to identify agonists and antagonists. The polypeptides, antibodies,
 CC agonists and antagonists (which are bacteriostatic) are used for the
 CC treatment and prevention of diseases such as otitis media in infants and
 CC children, pneumonia in elderlies, sinusitis, nosocomial infections and
 CC invasive diseases, and chronic otitis media with hearing loss. The
 CC polypeptides, agonists and antagonists are also used for screening of
 CC antibacterial drugs. The BAS034 products of the invention can be used
 CC screen for new antibacterial compounds that may target resistant
 CC bacteria.
 CC
 SQ Sequence 442 AA;
 Query Match 99.5%; Score 2348; DB 21; Length 442;
 Best Local Similarity 99.3%; Pred. No. 1.8e-217;
 Matches 439; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MKVSLSTLTLILSCFAILAIQQAQAVPNPVPFVDEVRSENDLQDNELPIDVQSATQSA 60
 CC |||||||
 Db 1 mkvslstltlilscfailaiqqaqavpnpvpfvdvrsndlgqdnelpidvqsatqsa 60
 CC |||||||
 QY 61 STDTANPLDEHEPELYTTALENTMLINCSALNODIMRLACYDTLVHGETPAVTKTKRSI 120
 CC |||||||
 Db 61 stdtanpldehepeelyttalenktmlincsalnqdimrlacydtlvhgecpavtktkrsi 120
 CC |||||||
 QY 121 RLDETIWQTIKGPQVYQETTDPIFLMGNEKGMLTKKDAKQLEYAAKQFTPLSLSPDL 180
 CC |||||||
 Db 121 rldetiwtikgkpqvvyqettdpiflmgnegkmltkkdaqlkleyaakqftplsld 180
 CC |||||||
 QY 181 RNNTPWSSRPHNPMYVLPFIMHGKPNRSPNTPSHEARQFTNPEFRAPELKFOVSVKVA 240
 CC |||||||
 Db 181 rnnptlwssrphnmpmyvlpfimhggkpnrsnptpsheakqftnpefrapelkfgsvkvka 240
 CC |||||||
 QY 241 AEDLWGTDSLWFGYTOQSHWQIFNGKNSRPFVRHVDYQPEIFLTQPVYSDLPWDGKVRMI 300
 CC |||||||
 Db 241 aedlwgtdslwfgytqgshwqifngknsrpfvrhdyqpeifltqpvysdlpwdgkvrmi 300
 CC |||||||
 QY 301 GMGAVHHSNGESAKLSRSNRRAYLMAGMEWKNLTVMPRIGRIFKEGSGSQPDNDPILD 360
 CC |||||||
 Db 301 gmgavhhsngesaklsrsnraylmagmewknlvtmprigrifkegsgsqpdndpild 360
 CC |||||||
 QY 361 YGYGQVDFRFLYQLENKSNISGTVRYNPRSGKQALQLDYVYPLGKISGYFQIFGQCSL 420
 CC |||||||
 Db 361 yygygvdfrflyqlenksnsgtvrynprrsgkqalqldyvypylgkigsgyfqiifggyqgs 420
 CC |||||||
 QY 421 IDYNHEATSEFGVGLMLNDWMGL 442
 CC |||||||
 Db 421 idynheatsfvgvglmndwmgl 442
 CC |||||||
 RESULT 4
 AAY85269
 ID AAY85269 standard; Protein; 442 AA.
 XX
 AC AAY85269;
 XX
 DT 29-JUN-2000 (first entry)
 XX
 DE BAS034 amino acid sequence #2.
 XX
 KW Moraxella catarrhalis infection; BAS034; diagnosis; staging;
 KW vaccine; bacteriostatic; treatment; prevention; otitis media; pneumonia;
 KW sinusitis; nosocomial infection; treatment; invasive disease; chronic otitis media;
 KW hearing loss; antibiotic drug.
 XX
 OS Moraxella catarrhalis.
 XX
 PN WO200015802-A1.
 XX
 PD 23-MAR-2000.
 XX
 PF 14-SEP-1999; 99WO-EP06781.
 XX
 PR 14-SEP-1998; 99WO-EP06781.

PR 14-SEP-1998; 98GB-00200002.
 XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX Ruelle J;
 XX WPI: 2000-271440/23.
 DR N-PSDB; AAA10701.
 XX Novel BASB034 polynucleotides and polypeptides from Moraxella
 PT catarrhalis used to prepare vaccines against bacterial infections
 XX
 PS Claim 3; Page 67; 106pp; English.
 CC This sequence represents a Moraxella catarrhalis BASB034 polypeptide from
 CC strain Mc2908. The invention relates to BASB034 polypeptides from
 CC M. catarrhalis strains Mc2931, Mc2908, Mc2913 and Mc2969. The BASB034
 CC polynucleotides and polypeptides may be employed as research reagents and
 CC material for the discovery of treatments and diagnostics for diseases,
 CC particularly human diseases. They are particularly used to diagnose and
 CC treat M. catarrhalis infections. They can be used for diagnosis of
 CC disease, staging of disease, or determining response of an infectious
 CC organism to drugs. The polynucleotides may be used as a source for
 CC hybridization probes, and for screening of genetic mutations, serotype,
 CC organism or strain identification, identification of mutations in BASB034
 CC sequences, and as components of arrays which are useful for diagnostic
 CC and prognostic purposes. The polypeptides can be used to produce
 CC antibodies. The polypeptides can also be used in vaccine formulations,
 CC and to identify agonists and antagonists. The polypeptides, antibodies,
 CC agonists and antagonists (which are bacteriostatic) are used for the
 CC treatment and prevention of diseases such as otitis media in infants and
 CC children, pneumonia in elderlies, sinusitis, nosocomial infections and
 CC invasive diseases, and chronic otitis media with hearing loss. The
 CC polypeptides, agonists and antagonists are also used for screening of
 CC antibacterial drugs. The BASB034 products of the invention can be used
 CC screen for new antibacterial compounds that may target resistant
 CC bacteria.
 XX
 SQ Sequence 442 AA;

Query Match 98.7%; Score 2330; DB 21; Length 442;
 Best Local Similarity 98.9%; Pred. No. 1e-215;
 Matches 437; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MKVSLSTLTSLSCFAITAIQAKAVNPVAFVDEVSENDLQDNELPDIQVQSATOSA 60
 Db 1 mkvslstlslscfaiatqakavnpvafvdevskndlgdngellvgvqatqsa 60
 Qy 61 STDTPANPLDEHEPELYTTALENKTMLINCSALNQDIMRLACYDTLVHGETPAVTKRSI 120
 Db 61 stdtanpldehepelyttalenktmlncsalnqdimrlacydtlvhgetpavtktrsi 120
 Qy 121 RLDEITWOTIKGKPOVVYQETDPIFLMGNEKGMTTKDAKOLEVAAKQFTPLSLSFOLD 180
 Db 121 rldetiwtikgkpoovyqetdplflmgnekgmttkdakolevaaqftplslsfold 180
 Qy 181 RNNTPLWSRRPHNMYVLPFIEMHGKPNRSPNTPSHEARQFTPNRAPELKQVSVKVA 240
 Db 181 rnntplwsrrphnmyvlpfihmgkpnrsptpshearqftpnrapelkfqvsvkva 240
 Qy 241 AEDLWGTDSDLWFGYQTQSHQOIFNGKNSRFRVHDYQPEIFLTPQVYSDLPWDGKVRMI 300
 Db 241 aedlwgtddslwfgyqtqshqoifngknsrfrvhdypeliftqpyvdsdlpwdgkvrmi 300
 Qy 301 GMGAVHSHNGESAKLSRSWKNAYLMAGHEWKNLTVMYPRIGWIRFKEGSGQDDNDPDIID 360
 Db 301 gmgavhshngesaklsrswknaylmagewknltvmprigwifrfkegsgqddndpdiid 360
 Qy 361 YYGVDVFLYQLENKNSISGTVRNPNSGKGLQDLYVPLGKISGYFOIFQGYGQSL 420
 Db 361 yygvdvfllyqlenknsgtvyrnpsrgkglqdyvypkigisgyfifqgygqsl 420

Qy 421 IDYNHEATSFVGVLMLNDMMGL 442
 Db 421 idynheatsfvgvlmlndwmg1 442
 RESULT 5
 AAAY75156
 ID AAY75156 standard; Protein: 370 AA.
 XX AC AAY75156;
 XX DT 21-MAR-2000 (first entry)
 XX DE Neisseria gonorrhoeae ORF 582 protein sequence SEQ ID NO:1786.
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.
 OS Neisseria gonorrhoeae.
 PN WO9957280-A2.
 XX PD 11-NOV-1999.
 XX PF 30-APR-1999; 99WO-US09346.
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX WPI: 2000-062150/05.
 DR N-PSDB; AAZ53918.
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX Claim 2; Page 903; 1453pp; English.
 XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAY75941
 CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX Sequence 370 AA;
 SQ

Query Match 34.8%; Score 822; DB 21; Length 370;
 Best Local Similarity 44.7%; Pred. No. 1.4e-70;
 Matches 163; Conservative 68; Mismatches 110; Indels 24; Gaps 8;
 Qy 87 INCSALNODIMRLACYDTLVHGETPAVI-----KTKRSIRLDETIWQTI-KGKPOVVYQE 140

```
Db 21 lqcaaltndvtrlacdrifaalqpsagqegqskavlnltetvrrssldkgeavivvek 80
QY 141 TTPDIFLMGNEKGMLTKKDAKQLEYAAKQFTPLSLGFDLDRNN-TPLWSSRPHPMYVLP 199
Db 81 ggdaI-----padsagetadlytplsmydlkndrlrgllgvrehmpmyimp 127
QY 200 IFMHGKPNRSPNTPSHEAR-QFTPNFRAPELKFQVSVKAAEDLWGTDSDLWFGYTOQ 258
Db 128 fwynnspnyapsptgtttvqekfgqktaetklqvsfkksiaenlfktradiwfgytr 187
QY 259 SHWQIEN-GENSRPFRVHDYQPEIFLTQPVYSDLPWDGKVRMIGMGAHVHNSGESAKLSR 317
Db 188 sdwqinyqrksapfrntdykpeifitqpvkadlpfggrlrmigafvfhqsgsqrsper 247
QY 318 SWNRAYLMAGMEKMLTVMPRIWGRIFKEGSGSQPDNDPILDYGYGVDYRFYQLENKS 377
Db 248 swnrlyamagmewgkltvprvvrarfdq-sgdk-ndnpladiadymgygdvklgyrindrq 305
QY 378 NISGTVRYNPRSGKGLQLDYVYPLGKIGSYGFQIFQGYGQSLIDYNHEATSFVGGLMLN 437
Db 306 nvysvlyrnpktygaaleaaytfpkkgklgvvrgfhgygeslidyhkhqngigiglmfn 365
QY 438 DWMGL 442
Db 366 dwdgi 370

RESULT 6
AAY75157
ID AAY75157 standard; Protein; 370 AA.
AC AAY75157;
DT 21-MAR-2000 (first entry)
XX Neisseria meningitidis ORF 582 protein sequence SEQ ID NO:1788.
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX Neisseria meningitidis.
OS Neisseria meningitidis.
XX WO9957280-A2.
XX 11-NOV-1999.
XX 30-APR-1999; 99WO-US09346.
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX WPI; 2000-062150/05.
DR N-PSDB; AA253919.
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
PS Claim 2; Page 904; 1453pp; English.
```

```
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 370 AA;
Query Match 34.5%; Score 815; DB 21; Length 370;
Best Local Similarity 44.7%; Pred. No. 6.7e-70;
Matches 163; Conservative 67; Mismatches 111; Indels 24; Gaps 8;
QY 87 INCSALNQDIMRLACDYTLVHGGETPAVI-----KTKRSIRLDETINQTI-KGKPPQVYQE 140
Db 21 lqcaaltndvtrlacdrifaalqpsagqegqskavlnltetvrrssldkgeavivvek 80
QY 141 TTPDIFLMGNEKGMLTKKDAKQLEYAAKQFTPLSLGFDLDRNN-TPLWSSRPHPMYVLP 199
Db 81 ggdaI-----padsagetadlytplsmydlkndrlrgllgvrehmpmyimp 127
QY 200 IFMHGKPNRSPNTPSHEAR-QFTPNFRAPELKFQVSVKAAEDLWGTDSDLWFGYTOQ 258
Db 128 lwynnspnyapsptgtttvqekfgqktaetklqvsfkksiaedlfktradiwfgytr 187
QY 259 SHWQIEN-GENSRPFRVHDYQPEIFLTQPVYSDLPWDGKVRMIGMGAHVHNSGESAKLSR 317
Db 188 sdwqinyqrksapfrntdykpeifitqpvkadlpfggrlrmigafvfhqsgsqrsper 247
QY 318 SWNRAYLMAGMEKMLTVMPRIWGRIFKEGSGSQPDNDPILDYGYGVDYRFYQLENKS 377
Db 248 swnrlyamagmewgkltvprvvrarfdq-sgdk-ndnpladiadymgygdvklgyrindrq 305
QY 378 NISGTVRYNPRSGKGLQLDYVYPLGKIGSYGFQIFQGYGQSLIDYNHEATSFVGGLMLN 437
Db 306 nvysvlyrnpktygaaleaaytfpkkgklgvvrgfhgygeslidyhkhqngigiglmfn 365
QY 438 DWMGL 442
Db 366 dwdgi 370

RESULT 7
AAY75158
ID AAY75158 standard; Protein; 370 AA.
AC AAY75158;
XX 21-MAR-2000 (first entry)
XX Neisseria meningitidis ORF 582 protein sequence SEQ ID NO:1790.
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX Neisseria meningitidis.
OS Neisseria meningitidis.
XX WO9957280-A2.
XX 11-NOV-1999.
XX 30-APR-1999; 99WO-US09346.
XX
```


Db	193	sdwqinyngqrksapfntdykpeifltqpvkaadlpfggrrlmlgaqfvhqsgngqsrpear	255			
QY	318	SWNRAYLMAGMEWKNLTVMPIRWGRIFKEGSGSQPDNDPILDYYGYGDVRFYQLLENKS	377			
Db	253	swriyamagmewgkltvprvvrfafdq-sgdk-ndnpdiadymgygdvklqyrlndrq	310			
QY	378	NISGTVRYNPRSRKGAQLQDYYVPLCKGKIGISGYFOITQGYGQSLIDYNHEATSFQVGLMLN	437			
Db	311	nvysviryngktgygaetaaytfpikgklkvrvrgfhgygeslidynhkqngigiglmfn	370			
QY	438	DWMGL 442				
Db	371	dlldgi 375				
RESULT 10						
AAW98871						
ID	AAW98871 standard; Protein; 355 AA.					
XX	AAW98871;					
DT	31-MAR-1999 (first entry)					
DE	H. pylori GHPO 1723 protein.					
XX	GHPO protein; Helicobacter infection; gastroduodenal disease; gastrit					
KW	peptic ulcer disease.					
OS	Helicobacter pylori.					
PN	W09843478-A1.					
XX	08-OCT-1998.					
XX	01-APR-1998; 98WO-US06371.					
PR	29-JUL-1997; 97US-0902615.					
PR	01-APR-1997; 97US-0833457.					
PR	24-JUN-1997; 97US-0881227.					
PA	(HUMA-) HUMAN GENOME SCI INC.					
PA	(INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.					
PI	Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;					
XX	WPI; 1998-542293/46.					
XX	N-PSDB; AAX14590.					
PT	New isolated Helicobacter polynucleotides - used to develop products					
PT	for the diagnosis, prevention and treatment of Helicobacter					
XX	infections and gastrointestinal diseases					
PS	Claim 8; Page 1976-1977; 2054pp; English.					
XX	This sequence represents a Helicobacter pylori GHPO protein of the					
CC	invention. The polypeptides can be used for preventing or treating					
CC	Helicobacter infections, and gastroduodenal diseases associated with					
CC	these infections, including acute, chronic, and atrophic gastritis, a					
CC	peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also					
CC	used for the production of antibodies. The products can also be used					
CC	detection and diagnosis.					
XX	Sequence 355 AA;					
SQ						
Query Match 10.4%; Score 246.5; DB 19; Length 355;						
Best Local Similarity 25.1%; Pred. No. 3.7e-15;						
Matches 82; Conservative 41; Mismatches 117; Indels 87; Gaps						
QY	157	KDKAKOLEYAAKQFTPLSLGSFSDLDRNNTPLWSSRPHNPYVLPFIWMHGKPNRSPNTPSHE	216			
Db	69	kkvinnmdvltqtlqfobfhsf-----tqifawyhonlhp-----	102			

QY 217 ARQFTPNRPAELKFOVSQVKKAAEDLWGTDSDLMFGYTOQSHQIFNGKNSRPRVRHD 276
 Db 103 ---yrnef-----kfqisfrvpvfrhllwtgtlylaytqtdwfgiypndpqsapmrmnn 154
 QY 277 YOPEIFLTOPVYSDLPWDCKV---RMIGMGAVHHSNG-ESAKLSRSWNRAYLMAGMEWKN 332
 Db 155 fmpeliyyvpi-nfkpfggkignfseiwgwhisngvggacypfnk-----egnpennq 209
 QY 333 LVMMPRI-----WGRIFKGGSGSQP-----DDNDP 357
 Db 210 fggqpvivkdyngqkdvrggcrsvsagrpvfrlvwkggikimvayvpyvpydqsnpn 269
 QY 358 ILDYGYGVDVRELY-----OLENKSNTSGTVRYNPRSGKALQLDYVYPLGKIGSYF 410
 Db 270 lidymgynakidyrrgrhhfelqlydiftqwyrd--rwbgafriygyrinpfvgiya 327
 QY 411 QIFOGYGOSLIDYNHEATSFVGGLMLN 437
 Db 328 qwfyngdglyeydvfsnrigvgirln 354

RESULT 11
 AAW20760
 ID AAW10960 standard; Protein; 356 AA.
 XX
 AC AAW10960;
 DT 08-JUN-1999 (first entry)
 DE H. pylori ORF 07ap80601_5083193_f3_8 cell envelope protein.
 KW Vaccine; probe; diagnostic; ORF; cell envelope protein;
 KW secreted protein; cellular protein.
 OS Helicobacter pylori.
 XX
 PN W09818323-AL.
 PD 07-MAY-1998.
 PF 28-OCT-1997; 97WO-US19575.
 XX
 PR 14-JUL-1997; 97US-0891928.
 PR 28-OCT-1996; 96US-0739150.
 PR 06-DEC-1996; 96US-0759739.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Alm RA, Smith D;
 XX
 DR WPI; 1998-271811/24.
 DR N-PSDB; AAX30427.
 XX
 PT Helicobacter pylori nucleic acids and proteins - used to develop
 PT products for the detection, prevention and treatment of H. pylori
 PT infections
 XX
 PS Claims 27, 31; Page 160-161; 279pp; English.
 XX
 CC Recombinant or substantially pure preparations of H. pylori polypeptides
 CC are disclosed, together with the nucleic acids encoding them. In all,
 CC 73 ORFs are shown. The proteins are variously cell envelope proteins,
 CC secreted proteins or other cellular proteins. Vaccines containing the
 CC nucleic acids or proteins are claimed, as are probes containing at least
 CC 8 nucleotides from the nucleic acid sequences. The vaccines are useful
 CC for treating or reducing the risk of H. pylori infections, and the
 CC probes can be used diagnostically for detecting the presence of
 CC Helicobacter in a sample. The products are also of use in screening
 CC for compounds having the ability to interfere with the H. pylori life
 CC cycle or to inhibit H. pylori infection.
 XX
 SQ Sequence 356 AA;

Query Match 10.1%; Score 239.5; DB 19; Length 356;
 Best Local Similarity 24.8%; Pred. No. 1.7e-14;
 Matches 82; Conservative 43; Mismatches 111; Indels 95; Gaps 12;
 QY 157 KDAKQLEVAQAQFTPLSLSFOLDNRNNTPLMSSRPHNPWYVLPIMHGKPNRSPNTPSHE 216
 Db 70 kkylnmmdylgtfipfyhsf-----tpifqwyhpninp----- 103
 QY 217 ARQFTPNRPAELKFOVSQVKKAAEDLWGTDSDLMFGYTOQSHQIFNGKNSRPRVRHD 276
 Db 104 ---yrnef-----kfqisfrvpvfrhllwtgtlylaytqtdwfgiypndpqsapmrmnn 155
 QY 277 YOPEIFLTOPVYSDLPWDCKV---RMIGMGAVHHSNG-ESAKLSRSWNRAYLMAGMEWKN 332
 Db 155 fmpeliyyvpi-nfkpfggkignfseiwgwhisngvggacypfnk-----egnpennq 209
 QY 333 LVMMPRI-----WGRIFKGGSGSQP-----DDNDP 357
 Db 210 fggqpvivkdyngqkdvrggcrsvsagrpvfrlvwkggikimvayvpyvpydqsnpn 269
 QY 358 ILDYGYGVDVRELY-----OLENKSNTSGTVRYNPRSGKALQLDYVYPLGKIGSYF 410
 Db 270 lidymgynakidyrrgrhhfelqlydiftqwyrd--rwbgafriygyrinpfvgiya 327
 QY 411 QIFOGYGOSLIDYNHEATSFVGGLMLN 437
 Db 328 qwfyngdglyeydvfsnrigvgirln 354

RESULT 12
 AAW20760
 ID AAW20760 standard; Protein; 253 AA.
 XX
 AC AAW20760;
 DT 15-JUL-1997 (first entry)
 DE H. pylori outer membrane protein, 07ap80601orf8.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope;
 KW outer membrane.
 OS Helicobacter pylori.
 XX
 PN W09640893-AL.
 XX
 PD 19-DEC-1996.
 XX
 PF 06-JUN-1996; 96WO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Berglindh OT, Smith D, Mellgaard BL;
 XX
 DR WPI; 1997-052306/05.
 DR N-PSDB; AAT68013.
 XX
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 PS Claim 56; Page 1172-1173; 1481pp; English.
 XX
 CC The present sequence is a Helicobacter pylori outer membrane protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds, useful
 CC as potential H. pylori life cycle activators or inhibitors. The genomic

CC sequence of *H. pylori* (ATCC 55679) was determined from overlapping
 CC contigs generated by mechanically shearing the bacterial DNA. The
 CC sequences were analysed for ORF of at least 180 nucleotides, and the
 CC predicted coding regions defined by computer evaluation. To identify
 CC likely *H. pylori* antigens for vaccine development, the amino acid
 CC sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from *H. pylori* by PCR amplification for recombinant polypeptide
 CC production, e.g. in *E. coli* hosts.

XX Sequence 253 AA;

Query Match 6.6%; Score 156.5; DB 18; Length 253;
 Best Local Similarity 25.6%; Pred. No. 1e-06;
 Matches 56; Conservative 26; Mismatches 72; Indels 65; Gaps 8;

QY 157 KDKAKOLEVAAKQFTPLSLFSLDRNTPMWSRPHPNMVLPDIFMHGKPNRSPNTPSHE 216
 DB 72 kkylnmdylygtfypfhsf-----tpifqwyhpnlnp----- 105

QY 217 ARQFTNEFRAPELKFOVSVKVAEDLNGTDSDLWFGYTOQSHMOIFENGKNSRPFVHD 276
 DB 106 ---yqrnef-----kqisfrvfpvfhilwktgtylaylqtnwrlqynopdsapmrmin 157

QY 277 YOPEIFLTPVYSDLPWDGKV---RMIGMGAVHHSNG-ESAKLSRSWNRAYLMAGMEWKN 332
 DB 158 fmpeliyyvpi-nfkpfqkignfseiwgwhisngvgagcyqfn----- 204

QY 333 LTVMPRIWRIFKEGSGSQ--PDDNPDIIDYGYGVDVRF 369
 DB 205 -----kegnpenqfpqpvivkdyngqkdvrw 231

RESULT 13

AAP70991 ID AAW70991 standard; Protein; 1686 AA.
 XX AC AAW70991;
 XX 19-OCT-1998 (first entry)
 XX Human class II P13 kinase-C2alpha.
 DE Human; class II phosphoinositide lipid kinase; P13 kinase;
 KW PIK3-C2alpha; class II; resistance; Wortmannin; LY294002.
 XX Homo sapiens.
 OS W09832864-A2.
 XX 30-JUL-1998.
 XX 27-JAN-1998; 98WO-GB00244.
 XX 28-JAN-1997; 97GB-0001652.
 XX (LUDW-) LUDWIG INST CANCER RES.
 XX Domin J, Waterfield MD;
 XX WPI; 1998-427960/36.
 DR N-PSDB; AAV42920.
 XX New nucleic acid encoding phosphoinositol kinase 3-C2 alpha or its
 PT fragments - useful for, e.g. treatment of tumour cells where
 PT phenotype is associated with expression of kinase
 XX Claim 3; Fig 1; 52pp; English.
 XX The present sequence represents a human class II phosphoinositide lipid
 CC (P13) kinase designated PIK3-C2alpha. It is characterised as a class II

CC kinase due to the presence of a conserved C2 domain found in murine and
 CC Drosophila class II P13 kinases, its apparent lack of a p85 binding site
 CC and a substrate affinity to inositol lipids Ptdins and Ptdins(4)P. The
 CC protein has resistance to P13 kinase inhibitors Wortmannin and LY294002.
 CC Antibodies against the protein (optionally humanised), are used to
 CC identify class II P13 kinases. Antisense sequences, antibodies or
 CC dominant negative mutants of the P13-C2alpha protein, are useful in human
 CC or veterinary medicine to block class II kinases. They can be used to
 CC treat tumour cells where the phenotype is associated with expression of
 CC P13-C2alpha protein.

XX Sequence 1686 AA;

Query Match 4.8%; Score 114; DB 19; Length 1686;
 Best Local Similarity 19.0%; Pred. No. 0.25;
 Matches 85; Conservative 68; Mismatches 154; Indels 140; Gaps 21;

QY 45 QDNELPIDV-----QSATOSASTDTANPLDE-----HEPELYTALENKTMLIN----- 88
 DB 522 eddetpvdlnkhlygiekpcckeamtrhpveellidsyhnqvelalqlenqhravdqvikav 581

QY 89 ---CSALNQDIMRLACYDILVHGETPAVVKTRKSIRLDETITQTIKGPQVVOETDPI 145
 DB 582 rkicsald-gvetial-----tesvkkikravnlpr-----ktadv 618

QY 146 FLMGNEKGMETKDKAKOLEVAAKQFTPLSLFSLDRNTPMWSRPHPNMV-----LP 199
 DB 619 slfge-----dtsrst-rgslnpenpvgvsinqltaa 651

QY 200 IF-----MHGKPNRSPNTPSHEARQFTNEFRAPELKFOVSVKVAEDL---WGTDSDLW 252
 DB 652 iydllrlhansgrsptdcagsskvskeawttteqlgtif---aahgissnwsnyeky 707

QY 253 FGYTQOSHMQIFNGKN-SRPF---RVHDIQPEIFLQ-----PV-YSDLPWDGKVMI 300
 DB 708 ylicslsh-----ngkdlfkpgskkvgtynkffylkwdehllfplqisqlplesvhlht 763

QY 301 GMGAVHHSNGESAKLSRSWN-----RAYLMAGMEWKNLTVMPRIWRIFK 345
 DB 764 lfgilngsgsspsdnkqkqpealqkvalplcdfrfltcg-----tkillyw----- 812

QY 346 EGSQSPDDNPDIIDYGYGVDVRFYQLENKSNISGTVRYNPRSGKALQLDVYVPLGKG 405
 DB 813 --tsstnsvpgtvkkgyvmerivlgvdfpsafdiitytpqvdrrsilqghnletlend 870

QY 406 ISGVFOIFQYGGOSLDYDYNHEATSFVG 432
 DB 871 ikg-----klidilhkdsllgl 887

RESULT 14

AAP70420 ID AAP70420 standard; protein; 800 AA.
 XX AC AAP70420;
 XX 20-JAN-1991 (first entry)
 XX Sequence encoded by cellulase gene derived from *Bacillus* sp. No. 1139.
 DE Enzyme; cellotriose; cellotetrose; hydrolysis.
 KW *Bacillus* sp. No. 1139.
 OS Key Location/Qualifiers
 FH Peptide 1..30
 FT Protein 31..800
 XX JP62232386-A.
 XX 12-OCT-1987.
 XX

Search completed: November 30, 2001, 14:17:00
Job time: 382 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:18:35 ; Search time 44.45 Seconds
(without alignments)
757.461 Million cell updates/sec

Title: US-09-787-083-8
Perfect score: 2363
Sequence: 1 MKVSLSTLTLSILPCFAILA.....YNHEATSFVGLMLNDMGL 442
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	815	34.5	382	2 H81195	phospholipase A1,
2	815	34.5	409	2 H81831	probable phospholip
3	373	15.8	286	2 B36971	outer membrane pho
4	366	15.5	329	2 D81279	phospholipase A1 (
5	358	15.2	289	2 A36971	outer membrane pho
6	357	15.1	289	1 PSECA1	phospholipase A1 (
7	357	15.1	289	2 E86069	outer membrane pho
8	321.5	13.6	289	2 C36971	outer membrane pho
9	246.5	10.4	355	2 C64582	phospholipase A1 p
10	240.5	10.2	355	2 H71930	probable phospholi
11	111.5	4.7	602	1 TVRTRR	protein kinase (EC
12	108.5	4.6	800	2 A29003	cellulase (EC 3.2.
13	108.5	4.6	822	2 JT0611	cellulase (EC 3.2.
14	107.5	4.5	824	3 JC7532	cellulase (EC 3.2.
15	106.5	4.5	901	2 T20122	hypothetical prote
16	100.5	4.3	660	2 E83656	methionyl-tRNA syn
17	100.5	4.3	783	2 A29003	cellulase (EC 3.2.
18	100.5	4.3	1658	2 JC5467	phosphoinositide 3
19	99.5	4.2	797	2 T46737	X-Pro dipeptidyl-p
20	99	4.2	5005	2 F82884	hypothetical prote
21	98.5	4.2	810	2 S49744	AMP deaminase (EC
22	98.5	4.2	838	2 A96557	probable receptor
23	96	4.1	791	2 H96839	hypothetical prote
24	96	4.1	4307	2 T20721	hypothetical prote
25	95.5	4.0	719	2 A42893	penicillin-binding
26	95	4.0	765	2 T35719	chitinase - Strept
27	95	4.0	888	2 S50801	AMP deaminase homo
28	94.5	4.0	397	2 A35136	cellulase (EC 3.2.
29	94.5	4.0	601	2 T26062	hypothetical prote

30	94.5	4.0	655	2 T26061	hypothetical prote
31	94.5	4.0	1310	2 T40135	probable involueme
32	94.5	4.0	1478	2 S20117	protein kinase BCK
33	93.5	4.0	564	2 T40777	ferric reductase t
34	93.5	4.0	719	2 S28031	penicillin-binding
35	93.5	4.0	719	2 S28033	penicillin-binding
36	93.5	4.0	857	1 A41369	S-receptor kinase
37	93	3.9	1078	2 T18352	gene 14 protein - Myc
38	92.5	3.9	486	2 S30959	penicillin-binding
39	92.5	3.9	719	2 S28034	penicillin-binding
40	92.5	3.9	719	2 S28032	penicillin-binding
41	92	3.9	324	2 B69521	hypothetical prote
42	92	3.9	470	2 T43675	cog-2 protein - Ca
43	91.5	3.9	796	2 JC7355	peroxisome prolife
44	91.5	3.9	1641	2 D82704	conserved hypothet
45	91	3.9	367	2 T24058	hypothetical prote

ALIGNMENTS

RESULT 1

E81195
phospholipase A1, probable NMB0464 [imported] - Neisseria meningitidis (strain MC58 s
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: E81195
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755
A:Accession: E81195
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-382 <TET>
A:Cross-references: GB:AE002403; GB:AE002098; NID:g7225688; PIDN:AAF40901.1; PID:g722
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0464

Query Match	34.5%	Score 815;	DB 2;	Length 382;
Best Local Similarity	44.7%	Pred. No. 3.5e-58;		
Matches 163;	Conservative 67;	Mismatches 111;	Indels 24;	Gaps 8;
QY	87	INCSALNQDIMRLACYDTLVHGETPAVI-----KTKRSIRLDETIWQTI-KGKPQVYVQE	140	
DB	33	LOCAALTNNVRLACYDRIFAALQPSAGQESKAVLNLTETVRSLLDKGEAVIVKE	92	
QY	141	TTDPIFLMGNEKMLTKDAKQLEYAAKQFTPLSLFSDLDLDRNN-TPLWSSRPINPMYVLP	199	
DB	93	GGDAL-----PADSAGETADITVPLSLMDLDRLLGLLGRHNPMLMP	139	
QY	200	IFMHGKNRSPNTPSH-EAKQFTPNFRAPFLQSVKVAEADLWGTDSLWFGYTOQ	258	
DB	140	LWYNNSPNAPGSPTRCTTVQEFQOKRAETKLVQVSKIAEDLFKTRADLWFGYTOR	199	
QY	259	SHWQIFN-GKNSRFRVHDYQPEIFITQPVYSDLPWDGKVRMIGMGAHVHNSGESAKLSR	317	
DB	200	SDWQIYNQGRKSAFFRNTDYKPEIFLTQPVKADLPFGGRLMLGAGFVHQSGQSPRES	259	
QY	318	SNWRYLMAGMEKNTVMPIRWGRIFKEGSGSQPDNDPILDYGYGVDRFELYOLENKS	377	
DB	260	SNWRYLMAGMEKNTVPIRWVWVAFDQ-SGDK-NDNPDIADYMGYGVDRFELYOLENKS	317	
QY	378	NSISGTVRYNPRSGKALQLDYVYPLGKIGSGYQIFQGYGQSLIDYNHEATSFVGLMLN	437	
DB	318	NYSVLRYNPKTGYGAIEAAYTPPIKGLKGVVGRFGHYGESLIDYNHKONGIGLMLFN	377	
QY	438	DMWGL	442	

[illegible]

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1999.
R;de Geus, P.; Verheij, H.M.; Riegman, N.H.; Hoekstra, W.P.M.; de Haas, G.H.
EMBO J. 3, 1799-1802, 1984
A>Title: The pro- and mature forms of the E. coli K-12 outer membrane phospholipase A.
A:Reference number: A00771; MUID:85003590
A:Accession: A00771
A:Molecule type: DNA
A:Residues: 'MTRQ', 34-289 <DEG>
R;Dekker, N.; Merck, K.; Tommassen, J.; Verheij, H.M.
Eur. J. Biochem. 232, 214-219, 1995
A>Title: In vitro folding of Escherichia coli outer-membrane phospholipase A.
A:Reference number: S66447; MUID:96048049
A:Accession: S66447
A>Status: preliminary
A:Molecule type: protein
A:Residues: 21-24 <DEK>
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: F65186
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-289 <BLAT>
A:Cross-references: GB:AE000458; GB:U00096; NID:g2367299; PIDN:AAC76824.1; PID:g2367373
A:Experimental source: strain K-12, substrain MG1655
C:Comment: This enzyme is tightly bound to the outer membrane of the cell.
C:Genetics:
A:Gene: pldA
A:Map position: 85 min
C:Superfamily: bacterial phospholipase A1
C:Keywords: carboxylic ester hydrolase; membrane bound
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-289/Product: phospholipase A1 #status predicted <MPT>

Query Match 15.1%; Score 357; DB 1; Length 289;
Best Local Similarity 39.3%; Pred. No. 2,5e-21;
Matches 84; Conservative 29; Mismatches 91; Indels 10; Gaps 4;

Qy 226 RAPELKFOVSVKKAAEDLGWTDSDLFYGYTQQSHWQIFNGKNRPFRVHDYQPEIFLTQ 285
I : I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I ::
Db 83 RKDEVQLSLAFPLWRGILGPNSVLGASVTQKSWOLSNSESSPPRETNYEPQLFLGF 142
I : I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I ::

Qy 286 PV-YSDLPWGCKVRWIGMGAVHHSGESAKLSRSNRAYLWAGMEWKNLTVMPRIWGRI 344
I : I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I ::
Db 143 ATDYRFAGW--TLRDVEVGYNHDSNGRSDPTSRSWNRLYTRLMAENGNNLVVEVKPWYVV- 199
I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I ::

Qy 345 KEGSQSDDNPDTLDYYGDVDFRYOLENKSNISGTVRYNPRSGKGALQLDYYVPLGK 404
I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I ::

Db 200 -----GNITDNDPDIKYGGYQLKIGYHL-GDAVLSAKGQYNWNTGYGGAEGLSYPITK 253
I : I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I ::

Qy 405 GISGYFOIFOGYGQSLIDYNHEATSFVGGLMLND 438
I : I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I ::

Db 254 HVRLYTQVSYGESLIDYNFNQTRVGVGMMLND 287
I : I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I :: I ::

RESULT 7
E86069
outer membrane phospholipase A [Imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-feb-2001 #text_change 31-Mar-2001
C:Accession: E86069
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E86069
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-289 <STO>

100

A;Experimental source: strain J99
C;Genetics:
A;gene: *pldA*

100

```
Query Match      10.2%; Score 240.5; DB 2; Length 355;
Best Local Similarity 25.1%; Pred. No. 8.9e-12;
Matches 83; Conservative 44; Mismatches 109; Indels 95; Gaps 13;

QY 157 KDKAKOLEYAAKQFTPLSLFSDLRNNTPLWSSRPHNPMYVLPFIEMHGKPNRSPNTPSHE 216
Db 69 KTYLNMDYLGTYFLPYHSF-----TFIFQWHPNINP----- 102

QY 217 AKQFTNEFRAPLEKQVSVKYKAAEDLWGTDSDLWFGYDQSHMQIFNGKNSRPRFVHD 276
Db 103 ---YORNEP-----KQISFRVPEVRHILTKGTLAYTOTNNFOYNDPOSAPHRMIN 154

QY 277 YQEIFLTPQVSDLPWDQKV-----RMIGCAVHHNG-ESAKLSRSWNR----- 321
Db 155 FMPELIYVPI-NFKPFGKGIGFSEIWIQWHSISGVGAQCYQPFNKEGPNENQFPQG 213

QY 322 -----AYLMAG-----MEWK--NLTVMPRIWGRIFKEGSGSQPDD- 354
Db 214 PVIVKDYNGQKDVWGGCRSVSAGNALCFVLVWEKGLKIMVAYWPYV-----PYDQ 265

QY 355 -NPDILDYGYGDVRFY-----OLENKSNIISGTVRYNPRSGKALQOLDYVYPLGKI 406
Db 266 SNQLIDYMGYGNKADYRGRHHFELQLYDFTQWRDYD--RWHGAFRLGYTYRINPFV 323

QY 407 SGYFQIQGYGQSLIDYNHEATSFYGLMLN 437
Db 324 GIYAQWFNGYGDGLYEYDFNSRIGVGIRLN 354

RESULT 11
TVTRR
N:Alternates: kinase (EC 2.7.1.37) raf - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 11-Jun-1999
C:Accession: B26126
R:ishikawa, F.; Takaku, F.; Nagao, M.; Sugimura, T.
Mol. Cell. Biol. 7, 1226-1232, 1987
A:Title: Rat c-raf oncogene activation by a rearrangement that produces a fused protein.
A:Reference number: A26126; MUID:87172791
A:Accession: B26126
A:Molecule type: mRNA
A:Residues: 1-602 <ISH>
A:Cross-references: GB:M15428; NID:G206546; PIDN:AAA42002.1; PID:G206547
C:Genetics:
A:Gene: raf
C:Superfamily: rat protein kinase raf; protein kinase homology
C:Keywords: ATP; autophosphorylation; phosphoprotein; phosphotransferase; serine/threonine
F:301-567/Domain: protein kinase homology <KIN>
F:309-317/Region: protein kinase ATP-binding motif
F:329/Active site: Lys #status predicted
F:453/Binding site: phosphate (Ser) (covalent) (by autophosphorylation) #status predicted

Query Match      4.7%; Score 111.5; DB 1; Length 602;
Best Local Similarity 19.3%; Pred. No. 0.53;
Matches 93; Conservative 71; Mismatches 174; Indels 145; Gaps 23;

QY 23 QAQAVPNPVAFVDEVRSENDLGDQNELPIDVQSATQSASTDTANPLDEHEPELYTTALEN 82
Db 81 ELEKLNQVVKLPEKKNKELETAQDNLGI-----QSQFTRAKEELEAEKRDILRT---N 131

QY 83 KTLMLNCSALNODIMRLACYDYLHGHETPAVIKTKRSIRLDETIWQ--TIKGRPOVYQOE 140
Db 132 ERUSQVEYLTEDVKRL--NEKLKESNT---TKGELQLQLDELQASDVTVKYREKRLQEQ 186

QY 141 -----TTDPIFLMGNKSG-----MLTKKDAKOLEYAAKQFTPLS 174
Db 187 KELLHNSWNLNLTAKTDELLALGREGKNEILELKCITLENKKEEDAIRSHSESAPSA 246

QY 175 LSFELDRNNTPLSSRPHNPMYVLPFIEMHGKPNRSPNTPSHEAKQFTPNEFRAPLEKFOV 234
Db 247 LS-SSPNLSPTGWSQPKTP-----VPAQRERAPCGSGTQEKNKIRPRGQRDSSYYWEI 298
```

```
QY 235 SVKVKAAEDLWGTDSDLWFGYDQSHW-----QIFNGKNSRPRFVHDYQPEI----- 281
Db 299 EASEVMLSTRIGSGS---FGTVYKWKWHDGVAVKILKVVDDPTPEQLQAFRNEAVLRKTR 355

QY 282 -----FLTQ-----PVYSDL--PWGKGVKRM-----GMCAVHH 307
Db 356 HVNILLFMGYMTKDLAIIVTQWCEGSSLYKHLHVQETKTFQMFQFQIIDIARTAGMDYLHA 415

QY 308 SNGESAKLSRSNRAVLMAGMEWK---NLTVMPRIWGRIFKEGSGSO----- 351
Db 416 KNIHRDMKS--NNIFLHEGLTVKIGDFGLATVKSRW-----SGSQVQEQPTGSLVLM 466

QY 352 -----PDNPDIL--DYGYGDVRFYQLENKSNISGTVRYNPRSGKALQOLDYVYPL 402
Db 467 APEVIRMODNNPFSFQSDVSYGIV--LYEL-----MTGELPYSHINNRDQI---IFMV 515

QY 403 GK 405
Db 516 GRG 518

RESULT 12
A29003
cellulase (EC 3.2.1.4), alkaline - Bacillus sp.
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus sp.
C:Date: 21-May-1988 #sequence_revision 21-May-1988 #text_change 21-Jul-2000
C:Accession: A29003
R:Fukumori, F.; Kudo, T.; Narahashi, Y.; Horikoshi, K.
J. Gen. Microbiol. 132, 2329-2335, 1986
A:Title: Molecular cloning and nucleotide sequence of the alkaline cellulase gene from
A:Reference number: A29003; MUID:87085443
A:Accession: A29003
A:Molecule type: DNA
A:Residues: 1-800 <FUK>
A:Cross-references: GB:D00066; GB:N00066; NID:G216223; PIDN:BAA00045.1; PID:G216224
A:Experimental source: strain 1139
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Superfamily: Bacillus sp. alkaline cellulase; Thermotoga xylanase A amino-terminal
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:585-726/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA>

Query Match      4.6%; Score 108.5; DB 2; Length 800;
Best Local Similarity 19.9%; Pred. No. 1.4;
Matches 96; Conservative 54; Mismatches 159; Indels 173; Gaps 25;

QY 8 LTLNILPCFALAIQQAQAVPNPVAFVDEVRSEND---LGQDNELPIDVQSATQSASTDT 64
Db 10 LISSILILVLLSL-----FPTALAEQVTRDNFKHLGNDNVKRPSEAGALQOEVDG 64

QY 65 ANPL-DEHEPELY-----TTALENKTMLINCSAL-----NQDIMRLACY-DTLVHGET 110
Db 65 QMTLVQHGKEIKQLRGMSTHGLQWFPPEILNDNAYKALANDWESNMIRLWVGVGNGVASN 124

QY 111 PAVIKTKRSIRLDETI-----NQF-----TKGKPOV 137
Db 125 PELIKSRVKGIDLATENDMYIVDVHVAHPGDPVPVAGAEDEFRRDIAALYPNNPHII 184

QY 138 YQETTPD-----IFLMGNEKMLTKKDAKOLEYAAKQFTPLSLSDLRN---NTPL 186
Db 185 YELANEPSSNNNGAGIPNNEEGWNAVK-----EYADPIVEMLRDSGNADNIIIVGSPN 239

QY 187 WSSRP-----HNPVYVLPFI--MHGKPNRS--PNTPSHEAKQFTNEFRAPLEKF 232
Db 240 WSORPDLAADNPIDDDHTMYTVHFYTGSHAASTESYPPETPSRNGVNMNTRYA----L 295

QY 233 QVSVKYKAAEDLWGT-----DSDLWFGYDQSHMQIFNGKNSRPRFVHDYQPEI----- 270
Db 296 ENGVAVFATE--NGTSQANGDGGPYDEADWIEFLNENNISWANWSLTN-KNEVSCAFT 352
```

Qy 271 PFRVH-----DYQPEIFLQPVYSDLPW--- 293
|| :
Db 353 PFLGKSNATSLDPCGQVWPPEELSLSGEYVRARIKGVNVEP-----IDRTKYTKVLWDFN 409
Qy 294 DGKVRMIGAVHHSNGES-----AKLSRSWNRAYLMAGWEKNTLWMPRIW 340
|| :
Db 410 DGTGKGFV-----NGDSPVEDVVIENEGALKLSGLDASNDVSEGNWYANARLSADGW 463
Qy 341 GR 342
Db 464 GK 465
RESULT 13
JT0611
cellulase (EC 3.2.1.4), alkaline - Bacillus sp. (strain KSM-64)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus sp.
C:Date: 09-Oct-1992 #sequence_revision 09-Oct-1992 #text_change 10-Dec-1999
C:Accession: JT0611
R:Sumitomo, N.; Ozaki, K.; Kawai, S.; Ito, S.
Biosci. Biotechnol. Biochem. 56, 872-877, 1992
A:Title: Nucleotide sequence of the gene for an alkaline endoglucanase from an alkalophilic
A:Reference number: JT0611; MUID:92305459
A:Accession: JT0611
A:Molecule type: DNA
A:Residues: 1-822 <SUM>
A:Cross-references: GB:M84963; NID:g289264; PIDN:AAA73189.1; PID:g289266
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Superfamily: Bacillus sp. alkaline cellulase; Thermotoga xylanase A amino-terminal rep
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:585/726/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA>
Query Match 4.6%; Score 108.5; DB 2; Length 822;
Best Local Similarity 19.9%; Pred. No. 1.5;
Matches 96; Conservative 54; Mismatches 159; Indels 173; Gaps 25;
Qy 8 LTLSLPCFAILAIQQAQVNPVAFVDEVRSEND---LGQDNELPIDVQSQASTDT 64
|| :
Db 10 LISSILILVLLSL-----FPTLAAEGNTREDNFKHLGNDNVKRPSEAGALQLEVDG 64
Qy 65 ANPL-DEHEPELY-----TTALENTMLINGCAL-----NQDIMRLACY-DTLVHGET 110
|| :
Db 65 QMTLVDQHGKIQLRGMSTHGLWFPPEILNDNAYKALANDNESNMIRLAMYVGENGYASN 124
Qy 111 PAVIKTKRSIRLDETI-----WQT-----IKGKQVY 137
|| :
Db 125 PELIKSRVIGIDLAIENDMYVVDVHVAHPCDPRDPVYAGAEDEFRIAALYPPNPHII 184
Qy 138 YQETTD-----IFLMGNEKGMTKKDAKQLEYAAKQFTPLSLSFOLDLN-----NTPIL 186
|| :
Db 185 YELANEPSSNNGGAGIPNNEEGNAVK-----EYADPIVEMLRSGNADNIIIVGSPN 239
Qy 187 WSSRP-----HNPWVLPF--MHGKPNRS--PNTPSHEAKQFTPNFEAPELKE 232
|| :
Db 240 WSRQPLADNPIDDHHTYVHTGSHAASTESYPETPNSEGNVMSNTRYA-----L 295
Qy 233 QVSVKVAADLWGT-----DSDLWFGYTOQ-----SHWQIFNGKNS-----R 270
|| :
Db 296 ENGVAVFATE--WGTSQLGDDGPPYDEADVWIEFLNENNISWANNSLTN-KNEVSGAFT 352
Qy 271 PFRVH-----DYQPEIFLQPVYSDLPW--- 293
|| :
Db 353 PFLGKSNATSLDPCGQVWPPEELSLSGEYVRARIKGVNVEP-----IDRTKYTKVLWDFN 409
Qy 294 DGKVRMIGAVHHSNGES-----AKLSRSWNRAYLMAGWEKNTLWMPRIW 340
|| :
Db 410 DGTGKGFV-----NGDSPVEDVVIENEGALKLSGLDASNDVSEGNWYANARLSADGW 463

Qy 341 GR 342
Db 464 GK 465
RESULT 14
JC7532
cellulase (EC 3.2.1.4), alkaline - Bacillus sp. (strain KSM-S237)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus sp. (strain KSM-S237)
C:Date: 31-Mar-2001 #sequence_revision 31-Mar-2001 #text_change 31-Mar-2001
C:Accession: JC7532; PC7107
R:Hakamada, Y.; Hatada, Y.; Koike, K.; Yoshimatsu, T.; Kawai, S.; Kobayashi, T.; Ito, T.
Biosci. Biotechnol. Biochem. 64, 2281-2289, 2000
A:Title: Deduced amino acid sequence and possible catalytic residues of a thermostabl
A:Reference number: JC7532; MUID:21036886
A:Accession: JC7532
A:Molecule type: DNA
A:Residues: 1-824 <HAK>
A:Cross-references: DDBJ:AB018420
A:Experimental source: strain KSM-S237
A:Accession: PC7107
A:Molecule type: protein
A:Residues: 31-50 <HA2>
C:Comment: This enzyme is a thermostable, alkaline enzyme used as effective detergent
C:Genetics:
A:Gene: Egl-237
C:Keywords: hydrolase; glycosidase
Query Match 4.5%; Score 107.5; DB 3; Length 824;
Best Local Similarity 19.7%; Pred. No. 1.8;
Matches 91; Conservative 56; Mismatches 147; Indels 167; Gaps 27;
Qy 30 PVAFVDEVRSEND---LGQDNELPIDVQSQASTDTANPL-DEHEPELY-----TT 78
|| :
Db 25 PAALAAEGNTREDNFKHLGNDNVKRPSEAGALQLEVDGQMTLVDQHGKIQLRGMSTH 84
Qy 79 ALENKTMILINGCAL-----NQDIMRLACYDTLVHGE-----TPAVIKTK-----RSI 120
|| :
Db 85 GLWFPPEILNDNAYKALSDNDWDSNMIRLAMY-----VGENGYATNPPELIKQKVIDGIELAI 140
Qy 121 RLDETI---WQT-----IKGKQVYVQETDP-----IF 146
|| :
Db 141 ENDMYVVDVHVAHPCDPRDPVYAGAKDFREIAALYPPNPHIYELANEPSSNNGGAG 200
Qy 147 LMGNEKGMTKKDAKQLEYAAKQFTPLSLSFOLDLN-----NTPLWSSRP-----H 192
Db 201 IPNNEEGKAVK-----EYADPIVEMLRSGNADNIIIVGSPNWSQRPDLAANDPIDDH 255
Qy 193 NPMVYLPFIMHGKPNRSPTPSHEAKQFTPNFEAPELK-----FQVSVKVAADLWGT 247
Db 256 HTMTVHTGSHAASTESYPSE-----TPNSERGNVMSNTRYALENGVAVFATE--WGT 308
Qy 248 -----DSDLWFGYTOQ-----SHWQIFNGKNS-----RPRV----- 274
Db 309 SQASGDGPPYDEADVWIEFLNENNISWANNSLTN-KNEVSGAFTPEELGKSNATNLDPG 367
Qy 275 --HDYQPEIFLQPVYSDLPW---OPV-----YSDLPW---DGKVRMIGAVHHSNGE 311
Db 368 PDHVAPEELSLSGEYVRARIKGVNVEPIDRTKYTKVLWDFNDGTKQ--GFCVNSDSNPK 425
Qy 312 SAKLSRSWNRAYLMAGME-----WKNLTVMPIWGR 342
Db 426 ELIAVDNENNTLKVSGLDVSDVSDGNFWANARLSANGWK 466
RESULT 15
T20122
hypothetical protein F25B3.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T20122; T21324

R. McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19225
A:Accession: T20122
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-901 <W12>
A:Cross-references: EMBL:Z70750; PIDN:CAA94750.1; GSPDB:GN00023; CESP:F25B3.1
A:Experimental source: clone C50F4
R:Gardner, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19406
A:Accession: T21324
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-901 <W12>
A:Cross-references: EMBL:Z70752; PIDN:CAA94758.1; GSPDB:GN00023; CESP:F25B3.1
A:Experimental source: clone F25B3
C:Genetics:
A:Gene: CESP:F25B3.1
A:Map position: 5
A:Introns: 25/3; 47/2; 199/1; 242/1; 315/2; 544/3; 600/1; 626/1; 646/1; 666/3; 666/3

	Query Match	4.5%; Score 106.5; DB 2;	Length 901;
	Best Local Similarity	19.5%; Pred. No.	2.5;
	Matches	74; Conservative	61; Mismatches 171; Indels 73; Gaps 16;
QY	31 VAFVDEVRENDLGO-----DNLPLDVQSATQASTDTANPLDBEHPELYTTALENK	83 : :: :	: :
Ddb	468 VAMITEIRNQKLEEAVDYRHYPDEVPVTPLASRNPALNQPOTDAADDDAETSNMRRFS	527 : :: :	: :
QY	84 TMLI-----NCNALQNQMRLACYYDTTLVHGGETPAVIKTKRISLDLETWQT	129 : : :	: :
Ddb	528 NYSITMVTPCVGAIRAASNRASSPKRDELQRARDLKIEKSTTAAATPNSRKASDEERRREE	587 : : :	: :
QY	130 I-----KKGPQQVVYOETDPILFMGNEKGMLTKKAOLEYAARQF----	TPLSLSFD 178 : :: :	: :
Ddb	588 VRRLMEKHQHPAIPSTSSPYTFTRIDGSNT--DURRIELDVHVKFKPDSPGLVRKQ	645 : :: :	: :
QY	179 LDRNNTPLMSRSSPHNPVMYLPIPMGHCKPNRSNP---TPS-----HEAQFTPTPEFAPEL	230 : : :	: :
Ddb	646 YDPNDT-----PH----VPAIGRGRTNGNRNDSSTPPSSASTFDRVKYR--GSMRSAEL	692 : : :	: :
QY	231 KFOVSVKVK-----AADBLMGTDSDLWFEGTYOQ--SIHQ-----IFNGKNRPFRVHDYOPE	280 : : :	: :
Ddb	693 KESLOLMMAKOYGMYGMGNDFESSODALATTPTKFSSOWEKDVDVEGTANELVLRIDERSID	752 : : :	: :
QY	281 IFLTQPVYSISDLPWDGKVRMIGMCVAHVHSNCEAKLSRSNWNAHYLAMAGMKWN-LTMVPRI	339 : :~ :	: :
Ddb	753 ITAAQADVIOD-----KIRETEVGSSSEEMLTASYLEYLTNERNTLVHRQEYNYNIETIROQ	807 : :~ :	: :
QY	340 WGRIFKEGS--GSQPDNDP	356 : :	: :
Ddb	808 TSELDOLGKIOINEVPDDPF	826 : :	: :

Search completed: November 30, 2001, 14:18:36
Job time: 258 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:17:48 ; Search time 36.79 Seconds
(without alignments)
270.358 Million cell updates/sec

Title: US-09-787-083-8
Perfect score: 2363
Sequence: 1 MKVSLSTLTLILPCFAILA.....YNHEATSGVGLMDNMGL 442

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100.5	4.3	1726	US-08-609-049A-30	Sequence 30, Appl
2	100.5	4.3	1726	US-09-170-996-30	Sequence 30, Appl
3	99.5	4.2	1658	US-08-609-049A-13	Sequence 13, Appl
4	99.5	4.2	1658	US-09-170-996-13	Sequence 13, Appl
5	95.5	4.0	682	US-08-481-435-6	Sequence 6, Appl
6	90.5	3.8	857	US-07-717-331F-2	Sequence 2, Appl
7	89.5	3.8	1024	US-09-091-117-5	Sequence 5, Appl
8	88.5	3.7	503	US-07-946-497-2	Sequence 2, Appl
9	88.5	3.7	503	US-08-483-322-2	Sequence 2, Appl
10	88.5	3.7	503	US-08-478-882-2	Sequence 2, Appl
11	87.5	3.7	666	US-08-961-083-2	Sequence 2, Appl
12	86	3.6	781	US-08-373-134D-2	Sequence 2, Appl
13	86	3.6	781	US-09-114-637-2	Sequence 2, Appl
14	86	3.6	985	PCT-US96-03916-6	Sequence 6, Appl
15	86	3.6	985	PCT-US96-03916-66	Sequence 66, Appl
16	85	3.6	816	US-07-731-157A-4	Sequence 4, Appl
17	85	3.6	816	US-08-229-444B-2	Sequence 2, Appl
18	85	3.6	816	US-08-541-780-4	Sequence 4, Appl
19	85	3.6	1627	US-07-665-792E-9	Sequence 9, Appl
20	84.5	3.6	774	US-08-902-632-2	Sequence 2, Appl
21	84.5	3.6	774	US-09-073-354-1	Sequence 1, Appl
22	84.5	3.6	774	US-08-656-005A-1	Sequence 1, Appl
23	84.5	3.6	774	US-09-073-259-1	Sequence 1, Appl
24	84.5	3.6	774	US-09-363-095-1	Sequence 1, Appl
25	84.5	3.6	774	US-09-418-027-1	Sequence 1, Appl
26	84	3.6	522	RE34606-6	Patent No. RE34,60
27	83.5	3.5	657	US-09-306-593-2	Sequence 2, Appl

28	83	3.5	1088	3	US-08-633-768A-1	Sequence 1, Appl
29	82.5	3.5	535	2	US-08-564-972-1	Sequence 1, Appl
30	82.5	3.5	535	4	US-09-171-969-9	Sequence 9, Appl
31	82.5	3.5	797	4	US-09-086-912-2	Sequence 2, Appl
32	82	3.5	320	2	US-08-245-511-4	Sequence 4, Appl
33	82	3.5	320	2	US-08-600-993A-4	Sequence 4, Appl
34	82	3.5	355	1	US-07-946-497-5	Sequence 5, Appl
35	82	3.5	355	1	US-08-483-322-5	Sequence 5, Appl
36	82	3.5	355	2	US-08-478-882-5	Sequence 5, Appl
37	82	3.5	1022	3	US-08-772-270A-2	Sequence 2, Appl
38	82	3.5	1381	4	US-09-540-245A-16	Sequence 16, Appl
39	81.5	3.4	663	1	US-08-441-139-7	Sequence 7, Appl
40	81.5	3.4	693	4	US-08-235-836C-68	Sequence 68, Appl
41	81.5	3.4	844	1	US-07-731-157A-6	Sequence 6, Appl
42	81.5	3.4	844	2	US-08-541-780-6	Sequence 6, Appl
43	81	3.4	334	2	US-08-359-850-4	Sequence 4, Appl
44	81	3.4	537	2	US-08-633-879C-2	Sequence 2, Appl
45	81	3.4	649	4	US-09-618-419-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-08-609-049A-30
; Sequence 30, Application US/08609049A
; Patent No. 5948664
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 5948664e1 PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,049A
; FILING DATE: 29-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-063700US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1726 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-609-049A-30

Query Match 4.3%; Score 100.5; DB 2; Length 1726;
Best Local Similarity 18.9%; Pred. No. 0.62;
Matches 86; Conservative 66; Mismatches 145; Indels 159; Gaps 21;
QY 45 QDNELPIDV-----QATOSASTDTANPLDE-----HEPELYTTALENKTMLIN----- 88
DB 563 EDDEAPVLDNKLQYIEKPYKVMTRHPVELDSDHYQVEL-ALQTEHQHRAVDQVKA 621

```

Qy 89 -----CSALNODIMLACVDTLVHGETP-----AVIKTKRSIRLDETIMQTIKGPQVYQOE 140
Db 622 VRKICSALED-----GVETPSVTEAVKKLRAVNLPR----- 652
Qy 141 TTDPIFLMGKMLTKDKAQLEYAAKQFTPLSLSFOLDNRNTPLMSSRRPHNPMYVLP- 199
Db 653 -----NKSADVT-----SLSGSDTRKSTKGSNLNPNPQVQVSM 686
Qy 200 -----IFMHGKPNR-----SPNTPSHEAKQFTPNFRAPELKFQVSVKVAEDL-- 244
Db 687 HLTRIIDYLLRLHANSRSCSTGCPGRSRNIKEAWTATE-----QLQFTVYAAHGISS 738
Qy 245 -WGTDSDLWFQYTOQSHWQIFNGKN-SRPF---RVHDYQPEIFLTQ-----PV-YSDL 291
Db 739 NWSNYEKYIILCSLH-----NGKDLFKPIQSKKVGTYKKNFFYLIKWDDELIIFFIQISQL 794
Qy 292 PWDGKVRMIGMCAVHHSNGESAKLSRSWN-----RAYLMAGMEWKNLTVM 336
Db 795 PLESVLHLTLFGLVNLQSSGSPDSNKQKQPEALGKVSILTFLDFKRFELTCG-----TKL 848
Qy 337 PRWGRIFKEGSGQPDNDPILDYGYGVDVRFYQLENKSNISGTVRYNPRSGKGALQL 396
Db 849 LYLW-----TSSHTNSIPGAIPKSVYMERIVLQVDFPSPAFDIITYTSPQIDRNIIQ 901
Qy 397 DYVYPLKGISGYFOIFQGYGQSLIDYNHEATSGV 432
Db 902 DKLETLESIDKG-----KLDDIHRDSSFG 927

```

```

RESULT 2
US-09-170-996-30
; Sequence 30, Application US/09170996
; Patent No. 6291220
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 6291220el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/170,996
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/609,049
; FILING DATE: 29-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-063700US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1726 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-170-996-30

```

```

Query Match 4.3%; Score 100.5; DB 4; Length 1726;
Best Local Similarity 18.9%; Pred. No. 0.62;
Matches 86; Conservative 66; Mismatches 145; Indels 159; Gaps 21;

Qy 45 QDNELPIDV-----QSATQASTDTANPLDE-----HEPELYTALENKTMNLIN----- 88
Db 563 EDDEAPVDNLKYLQIEKPYKEVMTRHVPVELLDSYHYQVEL-ALQTEHQHRAVDQVIKA 621
Qy 89 ----CSALNODIMLACVDTLVHGETP-----AVIKTKRSIRLDETIMQTIKGPQVYQOE 140
Db 622 VRKICSALED-----GVETPSVTEAVKKLRAVNLPR----- 652
Qy 141 TTDPIFLMGKMLTKDKAQLEYAAKQFTPLSLSFOLDNRNTPLMSSRRPHNPMYVLP- 199
Db 653 -----NKSADVT-----SLSGSDTRKSTKGSNLNPNPQVQVSM 686
Qy 200 -----IFMHGKPNR-----SPNTPSHEAKQFTPNFRAPELKFQVSVKVAEDL-- 244
Db 687 HLTRIIDYLLRLHANSRSCSTGCPGRSRNIKEAWTATE-----QLQFTVYAAHGISS 738
Qy 245 -WGTDSDLWFQYTOQSHWQIFNGKN-SRPF---RVHDYQPEIFLTQ-----PV-YSDL 291
Db 739 NWSNYEKYIILCSLH-----NGKDLFKPIQSKKVGTYKKNFFYLIKWDDELIIFFIQISQL 794
Qy 292 PWDGKVRMIGMCAVHHSNGESAKLSRSWN-----RAYLMAGMEWKNLTVM 336
Db 795 PLESVLHLTLFGLVNLQSSGSPDSNKQKQPEALGKVSILTFLDFKRFELTCG-----TKL 848
Qy 337 PRWGRIFKEGSGQPDNDPILDYGYGVDVRFYQLENKSNISGTVRYNPRSGKGALQL 396
Db 849 LYLW-----TSSHTNSIPGAIPKSVYMERIVLQVDFPSPAFDIITYTSPQIDRNIIQ 901
Qy 397 DYVYPLKGISGYFOIFQGYGQSLIDYNHEATSGV 432
Db 902 DKLETLESIDKG-----KLDDIHRDSSFG 927

```

```

RESULT 3
US-08-609-049A-13
; Sequence 13, Application US/08609049A
; Patent No. 5948664
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 5948664el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,049A
; FILING DATE: 29-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-063700US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 1658 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-609-049A-13

Query Match 4.2%; Score 99.5; DB 2; Length 1658;
Best Local Similarity 18.9%; Pred. No. 0.74;
Matches 101; Conservative 79; Mismatches 194; Indels 159; Gaps 23;

QY 5 LSTLTLSILPCFALLAIQ---QAQAVNPVAFVDEVR-----ENDLGQDNE 48
Db 381 LSPVTVORNCGENASVKVSTIEGLQLPVTFTCDVSVTEIIMQALCWVHDLNQ--- 437
QY 49 LPIDVQS-----ATQASASTDTANPLDEHE--PELYTTALENKTMLINCSALNODIMRLAC 101
Db 438 --VDVGSYILKVGQEEVLQNNHCLGSHQIQRKWDTEIKLQLLTLSAMCONLARTAE 495
QY 102 YDLVHGETPAVTKRSIRLDETIWQTIKGPQVVYQETDPF----- 146
Db 496 DD-----EAP-----VDLNKYLQIEKPYKEVMIRHPVEELDSYHYQVELALQTN 542
QY 147 -----LMGNEKGMTTKDAKOLEYAA-----KQFTPLSLSFDDLDRNN 183
Db 543 QHRAVDQVIKAVRKICSALDGVETPSVTEA-VKKLRVNLPRNKSADVTSLSGSDTRKN 601
QY 184 TPLWSSRPHPMVLP-----IFMHGKPNR-----SPNTPSHEAKOFTNEFRAPE 229
Db 602 STGSLNPENPVQVMDHLTTAIYDLRLHANSRSTGCGPRGSRNIKEAWTATE----- 656
QY 230 LKFQVSVKKAEDL---WGTDSDLWFGYQQSHWQIFNGKN-SRPF---RVHDIQPEIF 282
Db 657 ---QLQFTVVAAGISSNWVSNEYKYLICLSLH-----NGKDLFKPIQSKKVGTYKNFFY 709
QY 283 LTO-----PV-YSDLPWDGKVRMIGMGAHVHNSGESAKLSRSWN----- 320
Db 710 LKRWDELIIFPIQISQLPLESVLHLTLFGLVNLQSSGSSPDSNKQKRGKPEALGVSLTLFD 769
QY 321 -RAYLMAGMEKWLTVMPRIWGRIFKEGSGSQPDNDPILDYGYGVDVRFYOLENKSNI 379
Db 770 FKRLTCG-----TKLLYLW-----TSSHTNSIPGAIPKKSVMYMERIVLQVDFPSPA 816
QY 380 SGTVRYNPRSGKALQDLYVYPLGKGISGYFQIFQGYGQSLIDYNHEATSFV 432
Db 817 FDIITYSPQIDRNIQDKLETLESIDK-----KLLDIHRDSSFGL 859

RESULT 4

US-09-170-996-13
; Sequence 13, Application US/09170996
; Patent No. 6291220
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Moiz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 6291220el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/170,996

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/609,049
; FILING DATE: 29-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-0637000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1658 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-170-996-13

Query Match 4.2%; Score 99.5; DB 4; Length 1658;
Best Local Similarity 18.9%; Pred. No. 0.74;
Matches 101; Conservative 79; Mismatches 194; Indels 159; Gaps 23;

QY 5 LSTLTLSILPCFALLAIQ---QAQAVNPVAFVDEVR-----ENDLGQDNE 48
Db 381 LSPVTVORNCGENASVKVSTIEGLQLPVTFTCDVSVTEIIMQALCWVHDLNQ--- 437
QY 49 LPIDVQS-----ATQASASTDTANPLDEHE--PELYTTALENKTMLINCSALNODIMRLAC 101
Db 438 --VDVGSYILKVGQEEVLQNNHCLGSHQIQRKWDTEIKLQLLTLSAMCONLARTAE 495
QY 102 YDLVHGETPAVTKRSIRLDETIWQTIKGPQVVYQETDPF----- 146
Db 496 DD-----EAP-----VDLNKYLQIEKPYKEVMIRHPVEELDSYHYQVELALQTN 542
QY 147 -----LMGNEKGMTTKDAKOLEYAA-----KQFTPLSLSFDDLDRNN 183
Db 543 QHRAVDQVIKAVRKICSALDGVETPSVTEA-VKKLRVNLPRNKSADVTSLSGSDTRKN 601
QY 184 TPLWSSRPHPMVLP-----IFMHGKPNR-----SPNTPSHEAKOFTNEFRAPE 229
Db 602 STGSLNPENPVQVMDHLTTAIYDLRLHANSRSTGCGPRGSRNIKEAWTATE----- 656
QY 230 LKFQVSVKKAEDL---WGTDSDLWFGYQQSHWQIFNGKN-SRPF---RVHDIQPEIF 282
Db 657 ---QLQFTVVAAGISSNWVSNEYKYLICLSLH-----NGKDLFKPIQSKKVGTYKNFFY 709
QY 283 LTO-----PV-YSDLPWDGKVRMIGMGAHVHNSGESAKLSRSWN----- 320
Db 710 LKRWDELIIFPIQISQLPLESVLHLTLFGLVNLQSSGSSPDSNKQKRGKPEALGVSLTLFD 769
QY 321 -RAYLMAGMEKWLTVMPRIWGRIFKEGSGSQPDNDPILDYGYGVDVRFYOLENKSNI 379
Db 770 FKRLTCG-----TKLLYLW-----TSSHTNSIPGAIPKKSVMYMERIVLQVDFPSPA 816
QY 380 SGTVRYNPRSGKALQDLYVYPLGKGISGYFQIFQGYGQSLIDYNHEATSFV 432
Db 817 FDIITYSPQIDRNIQDKLETLESIDK-----KLLDIHRDSSFGL 859

RESULT 5

US-08-481-435-6
; Sequence 6, Application US/08481435
; Patent No. 6027906
; GENERAL INFORMATION:
; APPLICANT: Balganes, Tanjore S
; APPLICANT: Town, Christine
; TITLE OF INVENTION: No. 6027906el Polypeptides
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:

ADDRESSEE: White & Case
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2787
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/481.435
FILING DATE: 10-JUL-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IN 580/MAS/94
FILING DATE: 01-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 9404072-2
FILING DATE: 24-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sterner, Richard J.
REGISTRATION NUMBER: 35,372
REFERENCE/DOCKET NUMBER: 1103326-151
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 819-8783
TELEFAX: (212) 354-8113
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 682 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-481-435-6

Query Match 4.0%; Score 95.5; DB 3; Length 682;
Best Local Similarity 19.9%; Pred. No. 0.45;
Matches 89; Conservative 69; Mismatches 186; Indels 103; Gaps 24;

QY 18 ILAQQAQAVNPVAFVDEVRSENDL-GQDNELPDIQVSAQOS-----ASTDTTANP 67
DB 10 LVATTSSKIYNKKNQLIADGSEVRVNAQANDIPTDLVKAIVSIEDHRRFFDHRGIDTIRI 69
QY 68 LDEHEPELYTVALENTMLNCSALNQDIMRLACYDTLVHGETPAVIKRSIRLDETIW 127
DB 70 LGAFRLNQSLOGG-----SALTQQLIKLTYSTSTSDOTIS-----RKAQEAW 115
QY 128 QTIKGPQVYQET-----TDFIFLMGNEKGMILTK-----DKAKOLE-----YA 166
DB 116 LAIQLEQATKQEIITYINKVYMSNGYGMQTAQNYGKDLNLSLPQLALLAGMPQA 175
QY 167 AKQFTPLS-LSFDDLDRNTPLWSSRPHPMYVLPITFMHGKPNRSPNTPSHEAKQFTPNPF 225
DB 176 PNOYDPYSHPEAAQDRRLNLV--SEMKNGYI-----SAEQYEKAVNTPTDGLQSLKSA 229
QY 226 RAPE-----LKFQVS-VKVKAAEDLWGTDSLWFGYTO--QSH-KQIFNGKNSRFRVHD 276
DB 230 NYPAYMDNLYKEVINQVEETGYNLLTGMVDYTNVQEAQKHLMDIYNTDEYVAYPDE 289
QY 277 YQPEIFLQPVYSDLPWDGKRMIGMGAHVHRSN-----GESAKLSRSWNRNAYLMAGMEW 330
DB 290 LQ-----VASTIVDVS-NGKV-IAQLGARHQSNNVSGFNGQAVETNRD-----GSTM 335
QY 331 KNLT-VMPRIKRIKESGGSDPDNDILDYQ-----YGDVRFYOLENKS 377
DB 336 KPITDYAPALEYGVY-ESTATIVHDEP--YNYPGTNTPVYNMDRGYFGNITLQYALQOSR 392
QY 378 NISGTVRYN-----PRSGKALQLDY 398
DB 393 NYPAVETLNKVLGNRAKFTLNLGLIDY 419

RESULT 6

US-07-717-331F-2
; Sequence 2, Application US/07717331F
; Patent No. 5484905
; GENERAL INFORMATION:
; APPLICANT: June Nasrallah; Michael Nasrallah; and Joshua
; APPLICANT: Stein
; TITLE OF INVENTION: A Receptor Protein Kinase Gene
; TITLE OF INVENTION: Encoded At The Self-Incompatibility Locus
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yahwak & Associates
; STREET: 25 Skytop Drive
; CITY: Trumbull
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06611
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07717.331F
; FILING DATE: June 19th 1991
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: George M. Yahwak
; REGISTRATION NUMBER: 26,824
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203)268-1951
; TELEFAX: (203)268-1951
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 857 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-717-331F-2

Query Match 3.8%; Score 90.5; DB 1; Length 857;
Best Local Similarity 21.4%; Pred. No. 2.2;
Matches 76; Conservative 50; Mismatches 110; Indels 119; Gaps 23;

QY 105 LVHGETPAVIKRSIRLDETIWQIKGPKQVYQETTDPIFLMG-----NEK---GMLTK 157
DB 24 LIHPALSYINT-----LSSTESLTISSNKTLY---SPGSIFEVGFRTNSRWYLGMMYK 75
QY 158 K-DAKOLEYAAKQFTPLSLISFD-----LDRNNTPLWSSRPHNPMYVLPITFMHGK 205
DB 76 KVSDFTYVAVNRDNPISNAIGTLAGSNNLVLLDHSNKPVWNTN-----LTRGN 125
QY 206 PNRSP-----NTPSHEAKQFT-----PNEFRAPELKFQVSVKVKAAEDLW 245
DB 126 -ERSPPVAELLANGFVNRDSSNNDASEVLMQSFYDPTDITLPEMKLGYNLKT----- 177
QY 246 GTDSDLWFGYQQSHWQIFNGKNSRPRFRVHDYQ-----PEIFLQO-----PVYSDLPWDGK 296
DB 178 GLNRF-----TSWRSSDDPSGNGF---SYKLEQSLPEFLYLSRENFPNHRSGPWNG- 226
QY 297 VRMIGMGAVH-----HSNGESAKLSRSWNRAY-----LMAGMEWKNLTVMV--RI 339
DB 227 IRFSGIPEQDKLSYVYFNFIENNEEVAYFTMTNNSFTRLTISEGYFORLTWYPSIRI 286
QY 340 WGRIFKEGSGQDD-----NPDILDYGYGDVRFYLYLENKSNISGTVR--YNPRS 389
DB 287 WNRFWSSPVDQRCDYVIMCGP-----YAYCDV-----NTSPVCNCIOGFNPRN 329

RESULT 7

US-09-091-117-5

QY 260 HWQIFNGKN-----SRPRVHYOPEIFLTPVYSDLPW-----DGKVRMIGM 303
 Db 323 EQW---GKNPTPSDSHVTEGTASAHNNHPSQRMWTTQSQEDVSWTDFDFPISHPMGQG 379
 QY 304 AVHHSNGESA 313
 Db 380 HQTESKGHSS 389
 RESULT 9
 US-08-483-322-2
 ; Sequence 2, Application US/08483322
 ; Patent No. 5760178
 ; GENERAL INFORMATION:
 ; APPLICANT: HERRLICH, Peter
 ; APPLICANT: PONTA, Helmut
 ; APPLICANT: GUENTHER, Ursula
 ; APPLICANT: MATZKU, Siegfried
 ; APPLICANT: WENZL, Achim
 ; TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA
 ; TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
 ; TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington, D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/483.322
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/946,497
 ; FILING DATE: 09-NOV-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, Stephen A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 16915/145
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)672-5300
 ; TELEFAX: (202)672-5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 503 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-483-322-2

Query Match 3.7%; Score 88.5; DB 1; Length 503;
 Best Local Similarity 18.4%; Pred. No. 1.5;
 Matches 57; Conservative 32; Mismatches 96; Indels 125; Gaps 14;
 QY 40 ENDLGQDNELPIDVQSATQASDTDTANPLDHEPELYTTALENKTMLNCSALNODIMRL 99
 Db 169 QEDIDASNIIDEDVSS-----GSTIEKSTPGYILHLDLPT---SQPTGDRD----- 212
 QY 100 ACYDTLVHGETPAVIKTKRSI-----RLDETIWQTIKGPQVYQETTDPIFLMGNEKG 153
 Db 213 ---DAFFIGSLATATTPWWSAHTKQKQERTOWNPIHNSPEVLLQTT----- 258
 QY 154 MLTKKDAKQLEYAAKQFTPLSLSFDLDRNTPPL-----WSSRPHNPVYLPIFMHGKPNRS 209

Db 259 -----RMT-----DIDRNSTSAHGENWNTQEPQPPF----- 283
 QY 210 PNTPSHEAKQFTPNFEAPELKFAQSVKKAEDLW-----GTDSDLWFCYTOQS 259
 Db 284 -NNHEYQDEETPH-----ATSTTWADPNSTTEAAATQKEKWF-----EN 322
 QY 260 HWQIFNGKN-----SRPRVHYOPEIFLTPVYSDLPW-----DGKVRMIGM 303
 Db 323 EQW---GKNPTPSDSHVTEGTASAHNNHPSQRMWTTQSQEDVSWTDFDFPISHPMGQG 379
 QY 304 AVHHSNGESA 313
 Db 380 HQTESKGHSS 389
 RESULT 10
 US-08-478-882-2
 ; Sequence 2, Application US/08478882
 ; Patent No. 5885575
 ; GENERAL INFORMATION:
 ; APPLICANT: HERRLICH, Peter
 ; APPLICANT: PONTA, Helmut
 ; APPLICANT: GUENTHER, Ursula
 ; APPLICANT: MATZKU, Siegfried
 ; APPLICANT: WENZL, Achim
 ; TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA
 ; TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
 ; TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington, D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/478.882
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/946,497
 ; FILING DATE: 19921109
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, Stephen A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 16915/145
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)672-5300
 ; TELEFAX: (202)672-5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 503 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-478-882-2

Query Match 3.7%; Score 88.5; DB 2; Length 503;
 Best Local Similarity 18.4%; Pred. No. 1.5;
 Matches 57; Conservative 32; Mismatches 96; Indels 125; Gaps 14;
 QY 40 ENDLGQDNELPIDVQSATQASDTDTANPLDHEPELYTTALENKTMLNCSALNODIMRL 99
 Db 169 QEDIDASNIIDEDVSS-----GSTIEKSTPGYILHLDLPT---SQPTGDRD----- 212
 QY 100 ACYDTLVHGETPAVIKTKRSI-----RLDETIWQTIKGPQVYQETTDPIFLMGNEKG 153

Db 213 ---DAFFIGSTLTATTTPWVSAAHTKQONERTQWNP IHSNPEVLLQTTT----- 258
QY 154 MLTKKDAKQLEYAAKQFTPLSLSFOLDNRNTPL---WSSRPHNPMYVLP IEMHGKPNRS 209
Db 259 -----RMT-----DIDRNSTSAHGENWTQEPQPP----- 283
QY 210 PNTPSHAKQFTNEFRAPELKFQVSVKYKAAEDLW-----GTDSDLWFGYTOQS 259
Db 284 -NNHEYQDEETPH-----ATSTTWADPNSTTEEAATQKEKWF-----EN 322
QY 260 HQWIFNGKN-----SRPRVHDYQPEIFLTQPVYSDLPW-----DGKVRMIGMG 303
Db 323 EWO-----GKNPPTSPEDSHVTEGTASAHNNHPSQRMTTQSQEDVSWTDFDPIHSPMGOG 379
QY 304 AVHHSNGESA 313
Db 380 HQTESKGHSS 389

RESULT 11
US-08-961-083-2
; Sequence 2, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 666 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-2

Query Match 3.7%; Score 87.5; DB 4; Length 666;
Best Local Similarity 20.4%; Pred. No. 3;
Matches 87; Conservative 60; Mismatches 176; Indels 103; Gaps 24;
QY 42 DLG-----QNELPIDVQSATOS-----ASTDTANPLDEHEPELYTTALENKTMM 85
Db 12 DLGSRVNAQANDITDLVKAIVSIEDHRFDHGRIDIRILGAFLRNLQSNLOGG--- 69
QY 86 LINCALSALQDIMRLACYDTLVHGETPAVTKRSIRLDETIMQTIKGKQPVVYQET---- 141

Db 70 ---STLTQOLIKLYTFSTSDTOTIS-----RKAQEAWLAIQLEQKATKOEILTY 117
QY 142 TDPIFLMGNEKGLTK-----KDAKQLE-----YAAKQFTPLS-LSFDLDRNN 183
Db 118 INKVVMSNGNYGMQTAQNYGKDLNLSLPQALLAGMPQAPNQDYPYSHPEAAODRRN 177
QY 184 TPLWSSRPHNPMYVLP IEFMHGKPNRSPNTPPSHEAKOFTPNEFRAPE-----LKFEQVS-VK 237
Db 178 LVL--SEMKNQGYI-----SAEQYEKAVNTPITDGLQSLKSASNYPAYMDNYLKEVINQVE 231
QY 238 VKAAEDLWGTDSLWFGYTO--QSH-WQFNGKNSRPRFRVHDYQPEIFLTQPVYSDLPWD 294
Db 232 EETGYNLLTGMVYTVNVQEAQKHLMDIYNTDEYVAYPDDELQ-----VASTIVDVS-N 285
QY 295 GKVRMIGMGAVHHSN-----GESAKLSRSWNRAYLMAGMEWKNLT--VMPRI-----W 340
Db 286 GKV-IAQLGARHQSNNVSTGAINQAVETNRDW-----GSTMKPIITDYAPALEYGYVDST 337
QY 341 GRIFKEGSGQDDNPDI LDY-YCY-GDVRFYQLENKSNISCTVRYN-----PRSGKG 392
Db 338 ATIVHDEPNYPGTNTPVYNWMDRGYFGNITLQVALQOSRNVPAVETLKNVGLNRAKTEFN 397
QY 393 ALQLDY 398
Db 398 GLGIDY 403

RESULT 12
US-08-373-134D-2
; Sequence 2, Application US/08373134D
; Patent No. 5780296
; GENERAL INFORMATION:
; APPLICANT: Kmiec, Eric
; APPLICANT: Holloman, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS TO PROMOTE
; TITLE OF INVENTION: HOMOLOGOUS RECOMBINATION IN EUKARYOTIC CELLS AND ORGANISMS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/373,134D
; FILING DATE: January 17, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Friebe, Thomas E.
; REGISTRATION NUMBER: 29,258
; REFERENCE/DOCKET NUMBER: 7991-007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 781 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-373-134D-2

Query Match 3.6%; Score 86; DB 1; Length 781;
Best Local Similarity 20.7%; Pred. No. 5.5;

Matches 50; Conservative 37; Mismatches 122; Indels 32; Gaps 9;

Qy 39 SENDLGDQNELPIDVQS-ATOSASTDTANPLDEHEPELYTTALENKTMNLINCSALNQDIM 97
 Db 533 ARNDPGTSSQAPTSFASGCTGSALPDQPLAMDVASQTAFSGLLASIAPTLAEAVGAREL 592
 Qy 98 RLACYDVLHGETPAVITKRSIRLDETIWQTI-----KGKPVVYQE-----TTDPI 145
 Db 593 DSAC---ASNDVPLRLTEARTAQIGOT-WSNLINVRVFLSKTRARICMRDDQAPACEPV 647
 Qy 146 FLMGNEKMLTKKDAKOLEYAAKQFTPLSLF-FGLDRNNTPLWSSR---PHNPMYVLPI 200
 Db 648 RONTNQRGTASKSLMNTVRKAAVINPFGATMLDVGVDKSAALRQLRFRVITPRKAVHVL-- 705
 Qy 201 FMHGKPNRSPNTPSHEAKQFTNEFRAPELKQFVSVKAAEDLWGTDSDLWFGYTOQSH 260
 Db 706 -----NAYPSTVMH-AHATADSTPAPESQOQRAAERHPAEQEDADQDLFGEALQEH 758
 Qy 261 W 261
 Db 759 W 759

RESULT 13
 US-09-114-637-2
 ; Sequence 2, Application US/09114637
 ; Patent No. 5945339
 ; GENERAL INFORMATION:
 ; APPLICANT: Kmiec, Eric
 ; APPLICANT: Holloman, William
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS TO PROMOTE
 ; TITLE OF INVENTION: HOMOLOGOUS RECOMBINATION IN EUKARYOTIC CELLS AND ORGANISMS
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/114,637
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/373,134
 ; FILING DATE: January 17, 1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Friebe, Thomas E.
 ; REGISTRATION NUMBER: 29,258
 ; REFERENCE/DOCKET NUMBER: 7991-007
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-9741/8864
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 781 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-114-637-2

Query Match 3.6%; Score 86; DB 2; Length 781;
 Best Local Similarity 20.7%; Pred. No. 5.5;
 Matches 50; Conservative 37; Mismatches 122; Indels 32; Gaps 9;

Qy 39 SENDLGDQNELPIDVQS-ATOSASTDTANPLDEHEPELYTTALENKTMNLINCSALNQDIM 97
 Db 533 ARNDPGTSSQAPTSFASGCTGSALPDQPLAMDVASQTAFSGLLASIAPTLAEAVGAREL 592
 Qy 98 RLACYDVLHGETPAVITKRSIRLDETIWQTI-----KGKPVVYQE-----TTDPI 145
 Db 593 DSAC---ASNDVPLRLTEARTAQIGOT-WSNLINVRVFLSKTRARICMRDDQAPACEPV 647
 Qy 146 FLMGNEKMLTKKDAKOLEYAAKQFTPLSLF-FGLDRNNTPLWSSR---PHNPMYVLPI 200
 Db 648 RONTNQRGTASKSLMNTVRKAAVINPFGATMLDVGVDKSAALRQLRFRVITPRKAVHVL-- 705
 Qy 201 FMHGKPNRSPNTPSHEAKQFTNEFRAPELKQFVSVKAAEDLWGTDSDLWFGYTOQSH 260
 Db 706 -----NAYPSTVMH-AHATADSTPAPESQOQRAAERHPAEQEDADQDLFGEALQEH 758
 Qy 261 W 261
 Db 759 W 759

RESULT 14
 PCT-US96-03916-6
 ; Sequence 6, Application PC/TUS9603916
 ; GENERAL INFORMATION:
 ; APPLICANT: Wild, Martha A.
 ; APPLICANT: Cochran, Mark D.
 ; TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS
 ; TITLE OF INVENTION: AND USES THEREOF
 ; NUMBER OF SEQUENCES: 72
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US96/03916
 ; FILING DATE: 23-MAR-1995
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/126,597
 ; FILING DATE: 24-SEP-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 39116-A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 278-0400
 ; TELEFAX: (212) 391-0525
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 985 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US96-03916-6

Query Match 3.6%; Score 86; DB 5; Length 985;
 Best Local Similarity 19.4%; Pred. No. 8;
 Matches 73; Conservative 47; Mismatches 126; Indels 130; Gaps 20;

Qy 48 ELPIDVQSATOSASTDTANPLDEHEPELYTTALENKTMNLINCSALNQDIMRLACYDTLV 106
 Db 563 QTPSTVPKTTLSST-----EPAIFTRQSGTAFTQTSSAEPTMTQTSTH 613

Query Match 3.6%; Score 86; DB 5; Length 985;
Best Local Similarity 19.4%; Pred. No. 8;
Matches 73; Conservative 47; Mismatches 126; Indels 130; Gaps 20;

Search completed: November 30, 2001, 14:17:51
Job time: 228 sec

Search completed: November 30, 2001, 14:17:51
Job time: 228 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: November 30, 2001, 14:17:00 ; Search time 72.04 Seconds
(without alignments)
454.475 Million cell updates/sec
Title: US-09-787-083-8
Perfect score: 2363
Sequence: 1 MKVSLSTLTLILPCFAILA.....YNHEATSGVGLMLNDWML 442
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : A_Geneseq_1101.*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	2363	100.0	442	21	AA1985271
2	2354	99.6	442	21	AA1985268
3	2348	99.4	442	21	AA1985270
4	2339	99.0	442	21	AA1985269
5	822	34.8	370	21	AA1975156
6	815	34.5	370	21	AA1975157
7	815	34.5	370	21	AA1975158
8	815	34.5	374	21	AA1970629
9	812	34.4	375	21	AA1970628
10	246.5	10.4	355	19	AA198871
11	239.5	10.1	356	19	AA198960

12	156.5	6.6	253	18	AAW20760	H. pylori outer me
13	117	5.0	1686	19	AAW70991	Human class II P13
14	108.5	4.6	800	8	AAW70420	Sequence encoded b
15	108.5	4.6	822	13	AAW26021	Alkaline cellulase
16	107.5	4.5	824	21	AAW23180	Bacillus sp. KSM-S
17	107	4.5	157	18	AAW20538	H. pylori outer me
18	106	4.5	798	21	AAW40925	Human ORF689
19	106	4.5	798	21	AAW90225	Human PGC-1 protei
20	106	4.5	798	22	AAW83944	Peroxisome prolif
21	102.5	4.3	537	22	AAW91000	C glutamicum prote
22	100.5	4.3	1726	18	AAW38756	Phosphatidyl inos
23	98.5	4.2	467	20	AAW13378	Amino acid sequenc
24	98.5	4.2	467	21	AAW94860	Human protein clon
25	98.5	4.2	467	22	AAW88401	Human membrane or
26	98.5	4.2	467	22	AAW80246	Human PRO260 prote
27	95.5	4.0	502	22	AAW90031	C glutamicum prote
28	95.5	4.0	516	22	AAW78876	C. glutamicum SRT
29	95.5	4.0	682	17	AAW04359	S. pneumoniae penl
30	95	4.0	522	22	AAW95616	Human protein sequ
31	95	4.0	888	22	AAW70751	S cerevisiae apopt
32	94	4.0	372	16	AAW69607	Gonococcal porin-5
33	94	4.0	761	20	AAW99084	Non-B, non-C, non-
34	94	4.0	1227	22	AAW81501	S. epidermidis ope
35	94	4.0	1717	22	AAW20498	Ostrinia nubilalis
36	93.5	4.0	467	19	AAW75057	Human colon cancer
37	93	3.9	600	22	AAW75092	S receptor kinase
38	92.5	3.9	857	13	AAW29814	Cytadhesin P1. My
39	91	3.9	1627	16	AAW67538	Human protein sequ
40	90	3.8	522	22	AAW95513	Modified dipteria
41	90	3.8	560	11	AAW06815	T. matsutake pyran
42	90	3.8	564	21	AAW10457	Trichoderma deri
43	90	3.8	564	21	AAW81952	Human protein sequ
44	90	3.8	790	22	AAW93045	Mouse dephosphoryl
45	90	3.8	841	20	AAW24318	

ALIGNMENTS

RESULT	1
AAW85271	
ID	AAW85271 standard; Protein; 442 AA.
XX	
AC	AC
XX	AAW85271;
XX	
DT	29-JUN-2000 (first entry)
XX	
DE	BASB034 amino acid sequence #4.
XX	
KW	Moraxella catarrhalis infection; BASB034; diagnosis; staging;
KW	vacine; bacteriostatic; treatment; prevention; otitis media; pneumonia;
KW	sinusitis; nosocomial infection; invasive disease; chronic otitis media;
KW	hearing loss; antibacterial drug.
XX	
OS	Moraxella catarrhalis.
XX	
PN	WO200015802-A1.
XX	
PD	23-MAR-2000.
XX	
PF	14-SEP-1999; 99WO-EF06781.
XX	
PR	14-SEP-1998; 98GB-0020002.
XX	
PA	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX	
PI	Ruelle J;
XX	
XX	
DR	WPI; 2000-271440/23.
DR	N-PSDB; AAA10703.
XX	
XX	
PT	Novel BASB034 polynucleotides and polypeptides from Moraxella
PT	catarrhalis used to prepare vaccines against bacterial infections

XX PS Claim 3; Page 69; 106pp; English.
 XX CC This sequence represents a Moraxella catarrhalis BASB034 polypeptide from
 CC strain Mc2969. The invention relates to BASB034 polypeptides from
 CC M. catarrhalis strains Mc2931, Mc2908, Mc2913 and Mc2969. The BASB034
 CC polynucleotides and polypeptides may be employed as research reagents and
 CC material for the discovery of treatments and diagnostics for diseases,
 CC particularly human diseases. They are particularly used to diagnose and
 CC treat M. catarrhalis infections. They can be used for diagnosis of
 CC disease, staging of disease, or determining response of an infectious
 CC organism to drugs. The polynucleotides may be used as a source for
 CC hybridization probes, and for screening of genetic mutations, serotype,
 CC organism or strain identification, identification of mutations in BASB034
 CC sequences, and as components of arrays which are useful for diagnostic
 CC and prognostic purposes. The polypeptides can be used to produce
 CC antibodies. The polypeptides can also be used in vaccine formulations,
 CC and to identify agonists and antagonists. The polypeptides, antibodies,
 CC agonists and antagonists (which are bacteriostatic) are used for the
 CC treatment and prevention of diseases such as otitis media in infants and
 CC children, pneumonia in elderlies, sinusitis, nosocomial infections and
 CC invasive diseases, and chronic otitis media with hearing loss. The
 CC polypeptides, agonists and antagonists are also used for screening of
 CC antibacterial drugs. The BASB034 products of the invention can be used
 CC screen for new antibacterial compounds that may target resistant
 CC bacteria.
 XX SQ Sequence 442 AA;

Query Match 100.0%; Score 2363; DB 21; Length 442;
 Best Local Similarity 100.0%; Pred. No. 9.3e-219;
 Matches 442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MKVSLSTLTSLPCFALLAQQAQAVNPVAFVDEVSENDLGDNELPIDVQSATQSA 60
 Db 1 mkvslstltsilpcfallaiqqagavnpvafvdevrsendlgdnelpidvgsatqsa 60
 Qy 61 STDTPANPLDEHEPELYTTALENKTMLINCSALNODIMRLACYDTLVHGETPAVTKRSI 120
 Db 61 stdtanpldehepeylttalenktmlincsalnqdimrlacydtlvhgetpavtkrsl 120
 Qy 121 RLDETIWQITIKGPOVVQYQETDPIFLMGNEKGMTLTKDAKOLEYAAKQFTPLSLSFOLD 180
 Db 121 rldetiwtikgpkqvvyqettdpiflmgnekmttkdkakoleyaakqftplsfsfld 180
 Qy 181 RNTPLWSRRPHNPMVLPPIFMHGKPNRSPNTPSHEAKQFTPNFRAPLKFQSVKVA 240
 Db 181 rntplwsrrphnpmvlpipfmhgkpnrsntpsheakqftpnfrapelkfqsvkvka 240
 Qy 241 AEDLWGTSDSLWFGVTQSHWQIFNGKNSRPRFRVHDYQPEIFLTQPVYSDLPWDGKVRMI 300
 Db 241 aedlwgtstdslwfgvtqshwqifngknsrprfrvhdypelfltqpvysdlpwdgkvrm 300
 Qy 301 GMGAVHHSNGESAKLSRSWNRAYLMAGMEWKNLTVMYPRIGRIFKESGSGQPDNDPDI 360
 Db 301 gmgavhhsngesaklsrswnraylmagmewknltympriwgrifksgsgqpdndpdi 360
 Qy 361 YYGVDVRFLEKNSNIGTVRNPNSGKALQDLYVPLGKISGISYFQIFQYGGSL 420
 Db 361 yygvdvrflyqlenksnigtvrynprnsrgkalqldyvplgkigsgisfyfifqyggsl 420
 Qy 421 IDYNHEATSFVGGLMDNMGL 442
 Db 421 idynheatsfvgglmdnmgl 442

RESULT 2
 AAY85268
 ID AAY85268 standard; Protein; 442 AA.
 XX
 AC AAY85268;
 XX

DT 29-JUN-2000 (first entry)
 XX BASB034 amino acid sequence #1.
 XX Moraxella catarrhalis infection; BASB034; diagnosis; staging;
 KW vaccine; bacteriostatic; treatment; prevention; otitis media; pneumonia;
 KW sinusitis; nosocomial infection; invasive disease; chronic otitis media;
 KW hearing loss; antibacterial drug.
 XX OS Moraxella catarrhalis.
 XX WO200015802-A1.
 XX 23-MAR-2000.
 XX 14-SEP-1999; 99WO-EP06781.
 XX 14-SEP-1998; 98GB-0020002.
 XX (SMIK) SMITHLINE BEECHAM BIOLOGICALS.
 XX Ruelle J;
 XX WPI; 2000-271440/23.
 DR N-PSDB; AAA10700.
 XX Novel BASB034 polynucleotides and polypeptides from Moraxella
 PT catarrhalis used to prepare vaccines against bacterial infections -
 XX Claim 3; Fig 2; 106pp; English.

This sequence represents a Moraxella catarrhalis BASB034 polypeptide from strain Mc2931 (ATCC 43617). The invention relates to BASB034 polypeptides from M. catarrhalis strains Mc2931, Mc2908, Mc2913 and Mc2969. The BASB034 polynucleotides and polypeptides may be employed as research reagents and material for the discovery of treatments and diagnostics for diseases, particularly human diseases. They are particularly used to diagnose and treat M. catarrhalis infections. They can be used for diagnosis of disease, staging of disease, or determining response of an infectious organism to drugs. The polynucleotides may be used as a source for hybridization probes, and for screening of genetic mutations, serotype, organism or strain identification, identification of mutations in BASB034 sequences, and as components of arrays which are useful for diagnostic and prognostic purposes. The polypeptides can be used to produce antibodies. The polypeptides can also be used in vaccine formulations, and to identify agonists and antagonists. The polypeptides, antibodies, agonists and antagonists (which are bacteriostatic) are used for the treatment and prevention of diseases such as otitis media in infants and children, pneumonia in elderlies, sinusitis, nosocomial infections and invasive diseases, and chronic otitis media with hearing loss. The polypeptides, agonists and antagonists are also used for screening of antibacterial drugs. The BASB034 products of the invention can be used screen for new antibacterial compounds that may target resistant bacteria.

SQ Sequence 442 AA;

Query Match 99.6%; Score 2354; DB 21; Length 442;
 Best Local Similarity 99.5%; Pred. No. 6.8e-218;
 Matches 440; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKVSLSTLTSLPCFALLAQQAQAVNPVAFVDEVSENDLGDNELPIDVQSATQSA 60
 Db 1 mkvslstltsilpcfallaiqqagavnpvafvdevrsendlgdnelpidvgsatqsa 60
 Qy 61 STDTPANPLDEHEPELYTTALENKTMLINCSALNODIMRLACYDTLVHGETPAVTKRSI 120
 Db 61 stdtanpldehepeylttalenktmlincsalnqdimrlacydtlvhgetpavtkrsl 120
 Qy 121 RLDETIWQITIKGPOVVQYQETDPIFLMGNEKGMTLTKDAKOLEYAAKQFTPLSLSFOLD 180
 Db 121 rldetiwtikgpkqvvyqettdpiflmgnekmttkdkakoleyaakqftplsfsfld 180

QY 181 RNTPLWSSRRPHNMYVLP IFMHGKPNRSPNTPSHEAKQFTNFRAPELKFOVSVKVA 240
 |||||||
 Db 181 rntplwssrrphnmyvlp ifmhgkpnrsntpsheakqftnfrapelkfqsvkvka 240
 |||||||
 QY 241 AEDLWGTSDSLWFGYTOQSHWQIFNGKNSRPRVHDYQPEIFLTQPVYSDLPWDGKVRMI 300
 |||||||
 Db 241 aedlwgtstdslwfgytqshwqifngknsrprvrhdyqpeifltqpvysdlpwdgkvrm 300
 |||||||
 QY 301 GMGAVHHSNGESAKLSRSNWRAYLMAGMEWKNLTVMPTWGRIFKEGSGSQPDNDPILD 360
 |||||||
 Db 301 gmgavhhsngesaklsrswrnylmagmewknltvmptrwgrifkegsgsqpddndpild 360
 |||||||
 QY 361 YGYGDBVRFYQLENKSNISGTVRYNPRSGKALQLDVYVPLGKIGSYFOIFQGYGQSL 420
 |||||||
 Db 361 ygygdbvrflyqlenknsisgtvrynprsgkalgldvyvplgkigsgyfifqgygqsl 420
 |||||||
 QY 421 IDYNHEATSFVGVLMLNDWMGL 442
 |||||||
 Db 421 idynheatsfgvglmndwmgl 442
 |||||||
 RESULT 3
 ID AAY85270 standard; Protein; 442 AA.
 AC AAY85270;
 DT 29-JUN-2000 (first entry)
 XX BASH034 amino acid sequence #3.
 DE
 KW Moraxella catarrhalis infection; BASB034; diagnosis; staging;
 KW vaccine; bacteriostatic; treatment; prevention; otitis media; pneumonia;
 KW sinusitis; nosocomial infection; invasive disease; chronic otitis media;
 KW hearing loss; antibacterial drug.
 XX
 OS Moraxella catarrhalis.
 XX
 PN WO200015802-A1.
 XX
 PD 23-MAR-2000.
 XX
 PF 14-SEP-1999; 99WO-EP06781.
 XX
 PR 14-SEP-1998; 98GB-0020002.
 XX
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PI Ruelle J;
 XX
 DR WPI: 2000-271440/23.
 DR N-PSDB; AAA10702.
 XX
 PT Novel BASB034 polynucleotides and polypeptides from Moraxella
 catarrhalis used to prepare vaccines against bacterial infections
 XX
 PS Claim 3; Page 68; 106pp; English.
 XX
 CC This sequence represents a Moraxella catarrhalis BASB034 polypeptide from
 strain Mc2913. The invention relates to BASB034 polypeptides from
 M. catarrhalis strains Mc2913, Mc2908, and Mc2969. The BASB034
 CC polynucleotides and polypeptides may be employed as research reagents and
 CC material for the discovery of treatments and diagnostics for diseases,
 CC particularly human diseases. They are particularly used to diagnose and
 CC treat M. catarrhalis infections. They can be used for diagnosis of
 CC disease, staging of disease, or determining response of an infectious
 CC organism to drugs. The polynucleotides may be used as a source for
 CC hybridization probes, and for screening of genetic mutations, serotype,
 CC organism or strain identification, identification of mutations in BASB034
 CC sequences, and as components of arrays which are useful for diagnostic
 CC and prognostic purposes. The polypeptides can be used to produce
 CC antibodies. The polypeptides can also be used in vaccine formulations,

CC and to identify agonists and antagonists. The polypeptides, antibodies,
 CC agonists and antagonists (which are bacteriostatic) are used for the
 CC treatment and prevention of diseases such as otitis media in infants and
 CC children, pneumonia in elderlies, sinusitis, nosocomial infections and
 CC invasive diseases, and chronic otitis media with hearing loss. The
 CC polypeptides, agonists and antagonists are also used for screening of
 CC antibacterial drugs. The BASB034 products of the invention can be used
 CC screen for new antibacterial compounds that may target resistant
 CC bacteria.
 XX
 SQ Sequence 442 AA;
 Query Match 99.4%; Score 2348; DB 21; Length 442;
 Best Local Similarity 99.3%; Pred. No. 2.6e-217;
 Matches 439; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MKVSLSTLTLSPFAILAIQQAQAVPNPVARFVDSRSENDLGODNELPIDVQSQTSA 60
 |||||||
 Db 1 mkvslstlslscfaillaiqqakavpnpvarfvdsrsendlggdnepldvqsatqsa 60
 |||||||
 QY 61 STDTANPLDEHEPELYTTALENKTMLINGSALNQDLMRLACYDTLVHGETPAVTKRSI 120
 |||||||
 Db 61 stdtanpldehepeplyttalenktmlinsalnqdimrlacydtlvhgetpavtkkrsi 120
 |||||||
 QY 121 RLDETITWQTIKGPQVYQETDPIFLMGNEKGMLTCKDAKQLEYAAKQFTPLSLSFOLD 180
 |||||||
 Db 121 rldetiwtqtkgkpqvvyqettdpiflmgnekqmlckdakqleyaakqftplsfsld 180
 |||||||
 QY 181 RNTPLWSSRRPHNMYVLP IFMHGKPNRSPNTPSHEAKQFTNFRAPELKFOVSVKVA 240
 |||||||
 Db 181 rntplwssrrphnmyvlp ifmhgkpnrsntpsheakqftnfrapelkfqsvkvka 240
 |||||||
 QY 241 AEDLWGTSDSLWFGYTOQSHWQIFNGKNSRPRVHDYQPEIFLTQPVYSDLPWDGKVRMI 300
 |||||||
 Db 241 aedlwgtstdslwfgytqshwqifngknsrprvrhdyqpeifltqpvysdlpwdgkvrm 300
 |||||||
 QY 301 GMGAVHHSNGESAKLSRSNWRAYLMAGMEWKNLTVMPTWGRIFKEGSGSQPDNDPILD 360
 |||||||
 Db 301 gmgavhhsngesaklsrswrnylmagmewknltvmptrwgrifkegsgsqpddndpild 360
 |||||||
 QY 361 YGYGDBVRFYQLENKSNISGTVRYNPRSGKALQLDVYVPLGKIGSYFOIFQGYGQSL 420
 |||||||
 Db 361 ygygdbvrflyqlenknsisgtvrynprsgkalgldvyvplgkigsgyfifqgygqsl 420
 |||||||
 QY 421 IDYNHEATSFVGVLMLNDWMGL 442
 |||||||
 Db 421 idynheatsfgvglmndwmgl 442
 |||||||
 RESULT 4
 AAY85269
 ID AAY85269 standard; Protein; 442 AA.
 XX
 AC AAY85269;
 DT 29-JUN-2000 (first entry)
 XX
 DE BASB034 amino acid sequence #2.
 XX
 KW Moraxella catarrhalis infection; BASB034; diagnosis; staging;
 KW vaccine; bacteriostatic; treatment; prevention; otitis media; pneumonia;
 KW sinusitis; nosocomial infection; invasive disease; chronic otitis media;
 KW hearing loss; antibacterial drug.
 XX
 OS Moraxella catarrhalis.
 XX
 PN WO200015802-A1.
 XX
 PD 23-MAR-2000.
 XX
 PF 14-SEP-1999; 99WO-EP06781.
 XX

PR 14-SEP-1998; 98GB-00200002.
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX Ruelle J;
 XX WPI; 2000-271440/23.
 DR N-PSDB; AAA10701.
 XX Novel BASB034 polynucleotides and polypeptides from Moraxella
 PT catarrhalis used to prepare vaccines against bacterial infections -
 XX Claim 3; Page 67; 106pp; English.
 XX This sequence represents a Moraxella catarrhalis BASB034 polypeptide from
 CC strain Mc2908. The invention relates to BASB034 polypeptides from
 CC M. catarrhalis strains Mc2931, Mc2908, Mc2913 and Mc2969. The BASB034
 CC polynucleotides and polypeptides may be employed as research reagents and
 CC material for the discovery of treatments and diagnostics for diseases,
 CC particularly human diseases. They are particularly used to diagnose and
 CC treat M. catarrhalis infections. They can be used for diagnosis of
 CC disease, staging of disease, or determining response of an infectious
 CC organism to drugs. The polynucleotides may be used as a source for
 CC hybridization probes, and for screening of genetic mutations, serotype,
 CC organism or strain identification, identification of mutations in BASB034
 CC sequences, and as components of arrays which are useful for diagnostic
 CC and prognostic purposes. The polypeptides can be used to produce
 CC antibodies. The polypeptides can also be used in vaccine formulations,
 CC and to identify agonists and antagonists. The polypeptides, antibodies,
 CC agonists and antagonists (which are bacteriostatic) are used for the
 CC treatment and prevention of diseases such as otitis media in infants and
 CC children, pneumonia in elderlies, sinusitis, nosocomial infections and
 CC invasive diseases, and chronic otitis media with hearing loss. The
 CC polypeptides, agonists and antagonists are also used for screening of
 CC antibacterial drugs. The BASB034 products of the invention can be used
 CC screen for new antibacterial compounds that may target resistant
 CC bacteria.
 XX Sequence 442 AA;

Query Match 99.08; Score 2339; DB 21; Length 442;
 Best Local Similarity 99.18; Pred. No. 1.9e-216;
 Matches 438; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MKVSLSTLTSLPCFAILAIQQAQVNPVAFVDEVSENDLQDNELPDIQVQATQSA 60
 Db 1 mkvslstltlslpcfallaiqqagavnpvafvdevrskndlqgdnelllgvgsatqsa 60

Qy 61 STDTPANPLDEHEPELYTTALENKTMLINCSALNODIMRLACYDTLVHGETPAVINKTRSI 120
 Db 61 stdtanpldehepeylttalenktmlincsalnqdimrlacydtlvhgetpaviktrsi 120

Qy 121 RLDETINQTIKGVQVYQETDIFLGNKEKGLTKDKAQLEYAAKQFTPLSLFSLD 180
 Db 121 rldetinqtikgvqvyyetdtdiflmgnekglmtkdkaqleyaakqftplslfslfd 180

Qy 181 RNNTPLSSRRPHNPMYVLPFMHGKPNRSPNTPSHEAKOFTPNFRAPELKFOVSVKVA 240
 Db 181 rntplssrrphnpyvlpfmgkpnrsptpshearkoftpnfrapelfqvsvkva 240

Qy 241 AEDLWGTSDSLWFGYTOQSHQWQIFNGKNSRFRVHDYQPEIFLTQPVYSDLPWQKVRMI 300
 Db 241 aedlwgtsdslwfgytqshqwqifngknsrfrvhdypelfltqpvysdlpwqkvrm 300

Qy 301 GMGAVHNSGESAKLSRSWNPAYLWAGNWKNTVMPIRWIRIFKEGSGSQPDNDPDL 360
 Db 301 gmgavhnsgeaklsrswnpaylmagnewkntvmprirwifkegsgsqpdndpdl 360

Qy 361 YYGCDVRFYQLENKSNISGTQVYRNPGRSGKALQDLYVPLGKIGSYFOIFQYQOSL 420
 Db 361 yygcvdrfryqlenksnisgtvynprsgkalqldyvyplogkigsyfoifqyqgsl 420

Qy 421 IDYNHEATSFVGVLMLNDWMGL 442
 Db 421 idynheatsfvgvlnldwmgl 442

RESULT 5
 AAY75156
 ID AAY75156 standard; Protein; 370 AA.
 XX AAY75156;
 AC AAY75156;
 XX 21-MAR-2000 (first entry)
 DT
 XX Neisseria gonorrhoeae ORF 582 protein sequence SEQ ID NO:1786.
 DE
 XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.
 OS Neisseria gonorrhoeae.
 XX
 PN WO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI; 2000-062150/05.
 DR N-PSDB; AAZ53918.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 PS Claim 2; Page 903; 1453pp; English.
 XX
 CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SQ Sequence 370 AA;

Query Match 34.8%; Score 822; DB 21; Length 370;
 Best Local Similarity 44.7%; Pred. No. 1.6e-70;
 Matches 163; Conservative 68; Mismatches 110; Indels 24; Gaps 8;

Qy 87 INCSALNODIMRLACYDTLVHGETPAVI-----KTKRSIRLDETIWQTI-KGKPOVYQEQ 140

```
Db 21 lqcaaltdnvtlacydrifaalpsagqegeskavlnltetvrssldkgeavivvek 80
Qy 141 TTDPFLMGNEKMLTKKDAKOLEYAAKQFTPLSLFSLDRNN-TPLWSSRPHNPMYVLP 199
Db 81 ggdal-----padsagetadiytpislmldkndrlglgvrhmpmylmp 127
Qy 200 IFMHGKPNRSPNTPSH-EAKQFTPNFRAPELKFOVSVKAAEDLWGTDSLWFGYTQO 258
Db 128 fwnnspnyapsprgtvtqekfqqkraelqlvskfkskiaealnfktradiwfgytqr 187
Qy 259 SHWQIFN-GKNSRPRVHDYQPEIFLTQPVYSDLPWDGKVRMIGCAVHNSGESAKLSR 317
Db 188 sdwqlyngqrksapfrntdykpeifltqpvkadiplfggrlmlgafvghqsgqsrper 247
Qy 318 SNRRAYLMAGMEWKNLTVMPTWGRIFRFGKSGSQPDNDPDLIDYGYGDVRFYOLENKS 377
Db 248 swnrlyamagmewgklvprvvrfafq-sgdk-ndnpladiymgygdklqyrindrq 305
Qy 378 NISGTVRYNPRSGKALQLDYVYPLGKIGISGYFQIFQGYGOSLIDYNHEATSFVGLMLN 437
Db 306 nvyvslrynpktygaieaaytfpikgklgvrvrgfhgygeslidyhkhqngiglmfn 365
Qy 438 DMWGL 442
Db 366 dwdgi 370

RESULT 6
AA75157
ID AA75157 standard; Protein; 370 AA.
AC AA75157;
XX
XX
DT 21-MAR-2000 (first entry)
DE Neisseria meningitidis ORF 582 protein sequence SEQ ID NO:1788.
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
XX
PN WO9957280-A2.
PD 11-NOV-1999.
XX
XX
PF 30-APR-1999; 99WO-US09346.
XX
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizzi M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX
DR WPI: 2000-062150/05.
DR N-PSDB; AA253919.
XX
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
XX
XX
PS Claim 2; Page 904; 1453pp; English.
```

```
XX
CC AA253015 to AA254536, AA254577 to AA254615, and AA254253 to AA25941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia). To detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 370 AA;

Query Match 34.5%; Score 815; DB 21; Length 370;
Best Local Similarity 44.7%; Pred. No. 7.5e-70;
Matches 163; Conservative 67; Mismatches 111; Indels 24; Gaps 8;

Qy 87 INCSALNQDIMRLACYDTLVHGETPAVI-----KTKRSIRLDETIIWQTI-KGKPVVYQE 140
Db 21 lqcaaltdnvtlacydrifaalpsagqegeskavlnltetvrssldkgeavivvek 80
Qy 141 TTDPFLMGNEKMLTKKDAKOLEYAAKQFTPLSLFSLDRNN-TPLWSSRPHNPMYVLP 199
Db 81 ggdal-----padsagetadiytpislmldkndrlglgvrhmpmylmp 127
Qy 200 IFMHGKPNRSPNTPSH-EAKQFTPNFRAPELKFOVSVKAAEDLWGTDSLWFGYTQO 258
Db 128 lwnnspnyapsprgtvtqekfqqkraelqlvskfkskiaealnfktradiwfgytqr 187
Qy 259 SHWQIFN-GKNSRPRVHDYQPEIFLTQPVYSDLPWDGKVRMIGCAVHNSGESAKLSR 317
Db 188 sdwqlyngqrksapfrntdykpeifltqpvkadiplfggrlmlgafvghqsgqsrper 247
Qy 318 SNRRAYLMAGMEWKNLTVMPTWGRIFRFGKSGSQPDNDPDLIDYGYGDVRFYOLENKS 377
Db 248 swnrlyamagmewgklvprvvrfafq-sgdk-ndnpladiymgygdklqyrindrq 305
Qy 378 NISGTVRYNPRSGKALQLDYVYPLGKIGISGYFQIFQGYGOSLIDYNHEATSFVGLMLN 437
Db 306 nvyvslrynpktygaieaaytfpikgklgvrvrgfhgygeslidyhkhqngiglmfn 365
Qy 438 DMWGL 442
Db 366 dwdgi 370

RESULT 7
AA75158
ID AA75158 standard; Protein; 370 AA.
AC AA75158;
XX
XX
DT 21-MAR-2000 (first entry)
DE Neisseria meningitidis ORF 582 protein sequence SEQ ID NO:1790.
XX
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
XX
PN WO9957280-A2.
PD 11-NOV-1999.
XX
XX
PF 30-APR-1999; 99WO-US09346.
XX
```

RESULT	8	
AAAY70629		
ID	AAAY70629	standard; Protein; 374 AA.
XX		
AC	AAAY70629;	
XX		
DT	18-JUL-2000	(first entry)
XX		
DE	Neisseria meningitidis	serogroup B strain H44/76 BASB033 protein.
XX		
KW	BASB033;	diagnosis, prophylaxis; treatment; antibacterial; vaccine;
KW	Neisseria meningitidis	infection.
XX		
OS	Neisseria meningitidis.	
XX		
PN	WO200015801-A1.	
XX		
PD	23-MAR-2000.	
XX		
PF	09-SEP-1999;	99WO-EP06718.
XX		
PR	14-SEP-1998;	98GB-0020003.
XX		
PA	(SMIK)	SMITHKLINE BEECHAM BIOLOGICALS.
XX		
PI	Ruelle J;	
XX		
DR	WPI; 2000-271439/23.	
DR	N-PSDB; AA252134.	
XX		
PT	Isolated BASB033	polypeptides and polynucleotides of Neisseria
PT	meningitidis,	useful for diagnosis, prophylaxis and treatment of N.
PT	meningitidis	infection -
XX		
PS	Claim 4;	Page 59; 93pp; English.
XX		
CC	The present sequence	is a BASB033 protein from
CC	Neisseria meningitidis	serogroup B strain H44/76. The protein
CC	shows homology	to the Klebsiella pneumoniae outer membrane
CC	phospholipase A.	The present sequence is useful for diagnosis,
CC	prophylaxis and	treatment of N. meningitidis infection. It may also be
CC	used for the	discovery and development of antibacterial compounds and
CC	in vaccine	compositions.
XX		
SQ	Sequence	374 AA;
	Query Match	34.5%; Score 815; DB 21; Length 374;
	Best Local Similarity	44.7%; Pred. No. 7 6e-70;
	Matches 163;	Conservative 67; Mismatches 111; Indels 24; Gaps
QY	87	INCSALNQDLMRLACYDTLVHGETPAVI-----KTRSRILRDLFIWQTI-KGKPQVYQGE 140
DB	25	lqcaaltdnvtlacydrifaagqlpsagggeqgskavlnltetvrssldkgeavivvek 84
QY	141	TTDPFIPLMGNEKMLTKDAKOLEYAAKQFTPLSLFSLDFDLDRNN-TPLWSSRPHNPXVLP 199
DB	85	ggdai-----padsagetadiytlslmyldkndirgllgvrelnmpyImp 131
QY	200	IFVHGKPNRPNTPSH-EAKQFTPNFRAPELKFQVSVKAAEDLWGTDSDLWFGYTOQ 258
DB	132	lwynnspayagpaptgtvtqefgqakraetklqvsfkskaedlfktradiwfgytr 191
QY	259	SHWOIFN-GRNSRPRFVHDYQEIFETLPVYSDLPWDGKVRMTGMGAVHSHNGESAKLSR 317
DB	192	sdwgiyngqrksapfrntdykpeifitqpqkadiplfggrlrmigafvfhqsgsgspesr 251
QY	318	SWNRAYLMACMEWKNLTVMPRINWGRIFKEGSGSQDDNPDLIDYCYGVDVRFYLOLENKS 377
DB	252	swnrlyamngewgkltvlpfrvwrfafdq-sgdk-nhnpoladymgygavkqlqyrIndrq 309
QY	378	NISGTVRYNPRSGKALQLDYYVPLGKGISGYEQIFQGYGQSGLIDYNHEATISFGVGLMLN 437

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:26:28 ; Search time 77.9 Seconds
(without alignments)
829,941 Million cell updates/sec

Title: US-09-787-083-8
Perfect score: 2363
Sequence: 1 MKVSLSTLTLILPCFAILA.....YNHEATSFVGVLMLNDWMGL 442

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	815	34.5	382	2 Q9K0U7	Q9K0U7 neisseria m
2	815	34.5	409	2 Q9JT21	RT Q9JT21 neisseria m
3	372	15.7	306	2 Q9CL22	Q9CL22 pasteurella
4	366	15.5	329	2 Q9PMU8	Q9PMU8 campylobact
5	360.5	15.3	292	2 Q9SLD7	Q9SLD7 yersinia ps
6	360	15.2	289	2 Q9L6N9	Q9L6N9 salmonella
7	356.5	15.1	292	2 Q9Z4N8	Q9Z4N8 enterobacte
8	342.5	14.5	297	2 Q92349	Q92349 campylobact
9	246.5	10.4	355	2 Q92541	Q92541 helicobacte
10	240.5	10.2	355	2 Q92LX5	Q92LX5 helicobacte
11	123	5.2	278	2 Q9XB53	Q9XB53 erwinia car
12	117	5.0	1686	4 Q00443	Q00443 homo sapien
13	111.5	4.7	602	11 Q63485	Q63485 rattus norv
14	108.5	4.6	821	2 Q59241	Q59241 bacillus sp
15	107.5	4.5	824	2 Q9F216	Q9F216 bacillus sp
16	106.5	4.5	901	5 Q18749	Q18749 caenorhabdi
17	106	4.5	798	4 Q9UN32	Q9UN32 homo sapien
18	106	4.5	798	4 Q9UBK2	Q9UBK2 homo sapien
19	103	4.4	435	2 Q9X9C0	Q9X9C0 streptococc

ALIGNMENTS

RESULT 1
Q9K0U7 PRELIMINARY; PRT; 382 AA.
AC Q9K0U7;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PHOSPHOLIPASE A1, PUTATIVE.
GN NMB0464.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B., Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J., Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RA "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.";
RL Science 287:1809-1815(2000).
DR EMBL; AE002403; AAF40901.1; -.
DR TIGR; NMB0464; -.
DR InterPro; IPR003187; PLA1.
DR Pfam; PF02253; PLA1; 1.
KW Complete proteome.
SQ SEQUENCE 382 AA; 42714 MW; B468A802F062E836 CRC64;

Query Match 34.5%; Score 815; DB 2; Length 382;

Best Local Similarity 44.7%; Pred. No. 8.7e-63;

Matches 163; Conservative 67; Mismatches 111; Indels 24; Gaps 8;

QY 87 INCSALNQDIMRLACVDTLVHGETPAVI-----KTKRSRLDDETIWQTI-KGPKQVYVQE 140

DB 33 LQCAALTNDVTRLACVDRIFAALPSSAGOGESKAVLNLTETVRSSLDKGEAVIWEK 92

Query Match	34.5%	Score 815;	DB 2;	Length 409;
Best Local Similarity	44.7%;	Pred. No. 9.6e-63;		
Matches 163;	Conservative 67;	Mismatches 111;	Indels 24;	Gaps 8;
Qy	87	INCSALNODIMRLACACTYLVHGETPAVI-----KTKRSIRLDETIWQFI-KGKQVYVQE	140	
		: : : : : : : : : : : : : :		
Db	60	LQCAALTDNVTFLACYDRIFAAQLPSSAGQGQSKAVLNTETVRSLLDGEAVIVVEK	119	
		: : : : : : : : : : : : : :		
Qy	141	TTDPFIPLMGNEKGMILTKKDAQLEYAAKQFTPLSLSFOLDRNN-TPLWSSRPHNPMYVLP	199	
		: : : : : : : : : : : : : :		
Db	120	GGDAL-----PADSAGETADIYTPLSIMLYDLKNDLRLGLLVREHNPMYLMP	166	
		: : : : : : : : : : : : : :		
Qy	200	IFMHGKNRNSPNTPSH-EAKQFTNEFFRAPELKQTVSVKVVAAEDLWGTDSDLNFGYTTQ	258	
		: : : : : : : : :		
Db	167	LWYNNSNPAGSPTRGTVTGVEKEGQQRAETKLQVFSKIAEDLFTKRLADLNFYGTQR	226	
		: : : : : : : : :		

Qy	259	SHWOIFN-GKNSRPRVHDYQPEIFLTQPYYSDLPWDGKVMTGMGVAAHHSNGESAKLSR	317
Dd	227	SDWIYNQGRKSAPFRNTDYKPEIFLTQPVKADLPPFGRLRMILGAGVFHQSNQGSRPESR	286
Qy	318	SWNRAYLMAGNEWKNLVMPRIWGIRIEKSGSQDPDNDPDILDYGYGYDYRFYLQLENKS	377
Dd	287	SWNRIFYAMAGNEWKLVIPRVVVRAFDQ-SGDK-NONPDADIYMGYGDVKLOYRLMDRO	344
Qy	378	NISGTVRNPGRSGKALQLDDYVVLVGKGISGYFOIGYGOSLDIDYNHAEATSFVGGLMLN	437
Dd	345	NVYSVLRYNPKTGTGAERAYTFPIKGLKLGVVVRGFHYGESLDIDYHNKNGIGIGLMFN	404
Qy	438	DWMGL 442	
Dd	405	DLGGI 409	
RESULT	3		
ID	Q9CL22	PRELIMINARY; PRT; 306 AA.	
AC	O9CL22;		
DT	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DE	HYPOTHETICAL PROTEIN PM1426.		
GN	PM1426.		
OS	Pasteurella multocida.		
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;		
OC	Pasteurella		
OX	NCBI_Taxid=747;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=PM70;		
RX	MEDLINE=21145866; PubMed=11248100;		
RA	May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;		
RT	"Complete genomic sequence of Pasteurella multocida pm70.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 98:3460-3463(2001).		
DR	EMBL; AE006179; AAK03510.1; -		
DR	InterPro; IPR003187; PLAI.		
DR	Pfam; PF02253; PLA1; 1.		
KW	Hypothetical protein; Complete proteome.		
SQ	SEQUENCE 306 AA; 35580 MW; EAF3DE8CLC22B26E CRC64;		
Query Match	15.7%;	Score 372; DB 2; Length 306;	
Best Local Similarity	39.0%;	Pred. No. 2.5e-24;	
Matches	83; Conservative	40; Mismatches 82; Indels 8; Gaps	
Qy	229	ELKFQSVKYKAAEDLNGTDSLDWFQTQSHWQIFNGKNSRPRVRHDYQPEIFLTQPVY	288
Dd	97	EIKFKLSIALPLWGIIGNNSVLAASYTQSFWFSLSNVDSSPFRETNYEPQLFLAWKTQ	156
Qy	289	SDLPWGDKVMIGMGAVHHSNG--ESAKLSRSWNRAYLMAGMEWKNLTVMPRIWGIRIFE	346
Dd	157	YSLFPGTLQDVETGINHQSNGRDACKLSRNRLYVRSALKQNTVEIKPWRIPEK	216
Qy	347	GSGSQDPDNDILDYGYGDVRF-LYQLENKSNISGTVRYNPSRGKALQDYVYPLGKG	405
Dd	217	--AKNDNDNPITKYRGHFEDVALGYVYVYDHQHKLGS--HYNPTSKNGGLEASYSPITKN	271
Qy	406	ISGYFQJFYQGYQSLIDYNHEATSFVGGLMND 438	
Dd	272	IRFTQYNYNGESLDIQOIRIQIGIGISLN 304	
RESULT	4		
ID	Q9PMU8	PRELIMINARY; PRT; 329 AA.	
AC	Q9PMU8;		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		

[illegible]

```

Qy 345 KEGSQDDNDPILDYGYGDRFLYQLENKSNISGTVRYNPRSGKALQLDYVYPLGK 404
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 200 -----GSTDDNDPTKYGYQLKGYHL-GEAVLSAKGQYNWNTGYGAEVGLSYPYTK 253

Qy 405 GISGYFQIFQGYGSLIDYNHEATSGVGLMLND 438
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 254 HVRLYTVQYSGYGESLIDYNFQTRGVGVMLND 287

RESULT 7
Q924N8 PRELIMINARY; PRT; 292 AA.
AC Q924N8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE OUTER MEMBRANE PHOSPHOLIPASE A PRECURSOR (EC 3.1.1.32).
GN PLDA.
OS Enterobacter agglomerans.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94131966; PubMed=8300539;
RA Brok R.G., Brinkman E., van Bostel R., Bekkers A.C., Verheij H.M.,
RA Tommassen J.;
RT "Molecular characterization of enterobacterial plda genes encoding
RT outer membrane phospholipase A.";
RL J. Bacteriol. 176:861-870(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120145; PubMed=9921577;
RA Brok R.G., Boots A.P., Dekker N., Verheij H.M., Tommassen J.;
RT "Sequence comparison of outer membrane phospholipases A: implications
RT for structure and for the catalytic mechanism.";
RL Res. Microbiol. 149:703-710(1998).
DR EMBL; AF034414; AAD03498.1; -
DR InterPro; IPR003187; PLAI. 1
DR Pfam; PF02253; PLAI. 1
DR PRINTS; PR01486; PHPLIPASEAL.
KW SIGNAL; Hydrolase.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 292 OUTER MEMBRANE PHOSPHOLIPASE A.
SQ SEQUENCE 292 AA; 33719 MW; B75516D093B2BEEA CRC64;

Query Match 15.1%; Score 356.5; DB 2; Length 292;
Best Local Similarity 40.1%; Pred. No. 5.2e-23;
Matches 85; Conservative 31; Mismatches 87; Indels 9; Gaps 5;

Qy 229 ELKQVSKVKAEDLWCTDSDLWFGYTOQSHWQIFNCKNSRPFVRVHDYQPEIFLTQPV- 287
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 86 EVKFQSLGFFPIWIRGIADGNLSLGASYTQRSWQASNSDESSPFRETNYEPOIFLAWTD 145

Qy 288 YSDLPWGDKVRMIGMGAHHSGESAKLSRWNRAYLMAGMEWKNLTVMPTWGRIFRFG 347
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 146 YELAGW--TFREVEGYNHQSNADPTSRSDRYTTRMAQRGNLEIDLKPWYRIPESD 203

Qy 348 SGSPDDNDPILDYGYGDRFLYQLENKSNISGTVRYNPRSGKALQLDYVYPLGKI 406
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 204 S---KDDNPDKYNGYRLKVGALGESVFSIDG--RYNNTGYGGAEMGWSYPIRHHV 258

Qy 407 SGYFQIFQGYGSLIDYNHEATSGVGLMLND 438
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 259 RFTYQVSGYSGESMIDYNFQTRGVGVMLND 290

RESULT 8
Q92349 PRELIMINARY; PRT; 297 AA.
ID Q92349
AC Q92349;

```

```

DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PHOSPHOLIPASE A.
GN PLDA.
OS Campylobacter coli.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=195;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=UA585;
RC MEDLINE=97230284; PubMed=9119448;
RA Grant K.A., Belandia I., Dekker N., Richardson P.T., Park S.F.;
RT "Molecular characterization of plda, the structural gene for a
RT phospholipase A from Campylobacter coli, and its contribution to cell-
RT associated hemolysis.";
RL Infect. Immun. 65:1172-1180(1997).
DR EMBL; Y11031; CAA71915.1; -
DR InterPro; IPR003187; PLAI. 1
DR Pfam; PF02253; PLAI. 1
DR PRINTS; PR01486; PHPLIPASEAL.
SQ SEQUENCE 297 AA; 34998 MW; 04B54A7BCA3764CE CRC64;

Query Match 14.5%; Score 342.5; DB 2; Length 297;
Best Local Similarity 32.4%; Pred. No. 8.9e-22;
Matches 97; Conservative 37; Mismatches 108; Indels 57; Gaps 11;

Qy 178 DLDRNTPWLSRRPHNPMYVLPFMHGKPNRSPNT-----P 213
      || || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 16 DLKENNASLLSRKHETQ-----NTQKTPSTKEDFSRIALANYLGENSEFNPLGIS 66

Qy 214 SHEAKQETP-----NEFRAPELKQVSKVKAEDLWCTDSDLWFGYTOQSHWQIF 264
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67 SYKMNYFFPAYSGSLGGENRKTETKFKLSIKRKFEDLLGLGKGYVGYTQTSWQ-- 124

Qy 265 NGKNSRPFVRVHDYQPEIFLTQPV-YSDLPWGDKVRMIGMGAHHSGESAK--LSRSWNR 321
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 125 NYKHSSPFRETNYQPEFFVDIPLHFDYKFLNLR--VGLHESNGKGDENLESKNR 181

Qy 322 AYLMAEMWKNLTVMPTWGRIFRFGSGSQDDNDPILDYGYGDRFLYQLENKSNISG 381
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 182 IYASSVFLYQRFLEVPRIWYRI---PENSEDDNDNPEITHYMGNFIDIN-IGSLGNDYFNL 237

Qy 382 TVRYNP--RSCKGKALQLDYVYPL-CKGISGYFOIFQGYGSLIDYNHEATSGVGLMLN 437
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 238 MLRNLDLFDHDKGAVQVDIGDIFDNGIYWIYQYENGIGGSLIDINKRLQLRLTAFLIS 296

RESULT 9
O25241 PRELIMINARY; PRT; 355 AA.
AC O25241;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PHOSPHOLIPASE A1 PRECURSOR (DR-PHOSPHOLIPASE A).
GN HP0499.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,

```


RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen *Helicobacter*
RT *pylori*,"
RL Nature 388:539-547(1997).
DR EMBL: AE000564; AAD07564.1; --
DR TIGR: HP0499; --
DR InterPro: IPR003187; PLA1.
DR Pfam: PF02253; PLA1; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 355 AA; 42486 MW; 461960F15E65AB0A CRC64;

Query Match 10.4%; Score 246.5; DB 2; Length 355;
Best Local Similarity 25.1%; Pred. No. 2.7e-13;
Matches 82; Conservative 41; Mismatches 117; Indels 87; Gaps 11;

QY 157 KDKAKOLEYAAKQFTPLSLSFDLDRNNTPLWSSRPHPNMYVLPIFMHGKPNRSPNTPPSHE 216
DB 69 KKYLNMDYLGTYFLPFYHSF-----TFIQWYHPNINP----- 102
QY 217 AKOFTNEFRAPELKFOVSKVKAEDLWGTDSDLWFGYTQOSHWOIFNGKNSRPRFVHD 276
DB 103 ---YQRNEF-----KFOISFRVVPVFRHILWTKTGLYLAITQTNWFOIYNDDPOSAPMRMN 154
QY 277 YOPEIFLTPQVYSDLPWDGKV---RMIGMGAVHHSNG-ESAKLSRSWNRAYLMAGMEWKN 332
DB 155 FPELIYVYPI-NFKPFGKGKIGNFSEIWIGWOHISNGVGAQCYQPFNK-----EGNPENQ 209
QY 333 LTVMPRI-----WGRIFKEGSGSQP-----DDNPD 357
DB 210 FPGQPVIVDYNGQKDVWRGGRSVSAGQPVRLWKEGGLKIMVAYWPVYVDQSNEN 269
QY 358 ILDYGYGQDVRELY-----OLENKSNTSGTVRYNPRSGKALQLDYVYPLGKIGISGF 410
DB 270 LIDIMYGNAKIDYRGRHHELFQYDIFQYWRD--RWHGAFRLGTYRINPFVGIYA 327
QY 411 QIFQGYGQSLIDYNHEATSFVGGLMLN 437
DB 328 QWFNGYGDGLYEYDFNSRIGVGIRLN 354

RESULT 10
Q9ZLX5 PRELIMINARY; PRT; 355 AA.
AC Q9ZLX5;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PUTATIVE PHOSPHOLIPASE A1.
GN PLDA OR JHP0451.
OS *Helicobacter pylori* J99 (Campylobacter *pylori* J99).
OC Bacteria; Proteobacteria; epsilon subdivision; *Helicobacter* group;
OC *Helicobacter*.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen *Helicobacter pylori*,"
RL Nature 397:176-180(1999).
DR EMBL: AE001479; AAD06029.1; --
DR InterPro: IPR003187; PLA1.
DR Pfam: PF02253; PLA1; 1.
KW Complete proteome.
SQ SEQUENCE 355 AA; 42329 MW; B3CE9810EBA2FDC9 CRC64;

Query Match 10.2%; Score 240.5; DB 2; Length 355;
Best Local Similarity 25.1%; Pred. No. 8.9e-13;
Matches 83; Conservative 44; Mismatches 109; Indels 95; Gaps 13;

QY 157 KDKAKOLEYAAKQFTPLSLSFDLDRNNTPLWSSRPHPNMYVLPIFMHGKPNRSPNTPPSHE 216
DB 69 KKYLNMDYLGTYFLPFYHSF-----TFIQWYHPNINP----- 102
QY 217 AKOFTNEFRAPELKFOVSKVKAEDLWGTDSDLWFGYTQOSHWOIFNGKNSRPRFVHD 276
DB 103 ---YQRNEF-----KFOISFRVVPVFRHILWTKTGLYLAITQTNWFOIYNDDPOSAPMRMN 154
QY 277 YOPEIFLTPQVYSDLPWDGKV---RMIGMGAVHHSNG-ESAKLSRSWNR----- 321
DB 155 FPELIYVYPI-NFKPFGKGKIGNFSEIWIGWOHISNGVGAQCYQPFNKGNPENQFPQG 213
QY 322 -----AYLMAG-----MEWK--NLTVMPRWGRIFKEGSGSQPDD- 354
DB 214 PVIKDYNGQKDVWRGGRSVSAGNALCFVLWKEGGLKIMVAYWPV-----PYDQ 265
QY 355 -NPDILDYGYGQDVRELY-----OLENKSNTSGTVRYNPRSGKALQLDYVYPLGKGI 406
DB 266 SNQQLIDIMYGNAKIDYRGRHHELFQYDIFQYWRD--RWHGAFRLGTYRINPFV 323
QY 407 SGYFQIFQGYGQSLIDYNHEATSFVGGLMLN 437
DB 324 GIYAQWFNGYGDGLYEYDFNSRIGVGIRLN 354

RESULT 11
Q9XB53 PRELIMINARY; PRT; 278 AA.
AC Q9XB53;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE KDU1.
GN *Erwinia carotovora*.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC *Pectobacterium*.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 39048, GS101;
RX MEDLINE=98065591; PubMed=9402024;
RA McGowan S.J., Sebaihia M., O'Leary S., Hardie K.R., Williams P.,
RA Stewart G.S., Bycroft B.W., Salmund G.P.;
RT "Analysis of the carbenapem gene cluster of *Erwinia carotovora*:
RT definition of the antibiotic biosynthetic genes and evidence for a
RT novel beta-lactam resistance mechanism,"
RL Mol. Microbiol. 26:545-556(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 39048, GS101;
RX MEDLINE=98276484; PubMed=9614345;
RA McGowan S.J., Bycroft B.W., Salmund G.P.;
RT "Bacterial production of carbenapems and clavams: evolution of beta-
RT lactam antibiotic pathways,"
RN Trends Microbiol. 6:203-208(1998).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 39048, GS101;
RA McGowan S.J.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: U17274; AAD38237.1; --
SQ SEQUENCE 278 AA; 31341 MW; 63769F4F3550E1B7 CRC64;

Query Match 5.2%; Score 123; DB 2; Length 278;
Best Local Similarity 21.6%; Pred. No. 0.011;
Matches 74; Conservative 52; Mismatches 128; Indels 88; Gaps 16;

Db	9	LISSILLIVLLLSL-----PPTALAAEGNTRDNFXKHLGNDNVKRPSEAGALQLOEVDG	63
Qy	65	ANPL-DEHEPELY-----TTALENKMLINCALSAL-----NODIMRLACY-DTLVHGET	110
Db	64	QMTLVDGHGEKIQLRGSMSTHGLQWFFPEILNDNAYKALANDWESNMIRLAMYGENGYASN	123
Qy	111	PAVTKTKRSIRLDETI-----NQT-----IKGKPQGV	137
Db	124	PELKSRVIGIDLATENDMVIVDHHVHAPGDPDPVYAGADEDFRDIAALYPNPHII	183
Qy	138	YQETTPD-----IFLMGNEKMLTKDKAQLEYAAKQFPLSLSFDDLDRN-----NTPL	186
Db	184	YELANEPSSNNNGGAGIPNNEEGWNAVK-----EYADPIVEMLRDGSNADDNIIIVGSPN	238
Qy	187	WSSRP-----HNPMVYLPIF--MHGKPNRS--PNTPSHEAKQFPTNEFRAPELKF	232
Db	239	WSQAPDLAANPIDDDHHTMTYVHFYTGSHAATESTPTPNSRGNVMSNTRYA-----L	294
Qy	233	QVSVKVKAEDLMGT-----DSDLWFQYQQ-----SHQWIFNGKNS-----R	270
Db	295	ENGVAVFATE--NGTISOANGDGGPYDDEADVWIEFLNENNISWANSULTN-KNEVSGAFT	351
Qy	271	PFVRH-----DYPQEIFLTPQVYSDLPW---293	
Db	352	PFELGKSNATSLDPPGPDQVWVPELSLSGEYVRARIKGVNYPE--IDRTKYTKVLWDFN	408
Qy	294	DGKVRMTIGMAVHNSGES-----AKLSRSMNRYALMAGMEWKNTLTMVPRW	340
Db	409	DGTKQGEV-----NGDSPVEDVVIENEAGALKLSGLDASNDVSGENYWANARLSADGW	462
Qy	341	GR	342
Db	463	GK	464

```

RESULT 15
Q9F216 PRELIMINARY; PRRT; 824 AA.
ID Q9F216;
AC Q9F216;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE CELLULOSE.
DE Bacillus sp.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1409;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KSM-S237;
RX MEDLINE=21036886; PubMed=11193393;
RA Hakamada Y., Hatada Y., Koike K., Yoshimatsu T., Kawai S.,
RA Kobayashi T., Ito S.;
RT "deduced amino acid sequence and possible catalytic residues of a
RT thermostable, alkaline cellulase from an alkaliphilic Bacillus
RT strain.";
RL Biosci. Biotechnol. Biochem. 64:2281-2289(2000).
DR EMBL: AB018420; BAB19360.1; -.
DR InterPro: IPR001547; glyco_hydro_F5.
DR Pfam: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN.1.
SQ SEQUENCE 824 AA; 91564 MW; 65FA940FE1D729B9 CRC64;

```

```

Query Match      4.5%; Score 107.5; DB 2; Length 824;
Best Local Similarity 19.1%; Pred. No. 1.2;
Matches 91; Conservative 56; Mismatches 147; Indels 167; Gaps 27;

QY 30 PVAFVDEVRSEND-----LQGDNELPFDVGSATQSASTDTANPL-DEHEPELY-----TT 78
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25 PAALAAEGNTREDNFKHLGNNDVNVKRESEAGALQLQVEVDGQMTLDVDOGHEKIQLRGSMSTH 84
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 79 ALENTWMLNCSAL-----NODIMRLACVDTLVHCE-----TPAVIKTK-----RSI 120

```

```
Db 85 GLQWPEILNDNAYKALSNDSDNNIRLAMY-----VGENGYATNPelikQrVIDGIELAI 140
Qy 121 RLDETI---WQT-----IKGKQVVOYQETTD-----IF 146
Db 141 ENDMYVIVDWHVHAPGDRDPVYAGAKDFFREIAALYPNNPHIIYELANEPSSNNNGGAG 200
Qy 147 LMGNEKGLTKKDAKQLEYAAKQFTPLSLSFOLDN-----NTPLWSSRP-----H 192
Db 201 IPNNEEGKAVK-----EVADPIVEMLRKSGNADNIIIVGSPNWSQRPDLAADNPIDDH 255
Qy 193 NPMYVLPPIFMHGCKPNRSPNTPSHEAKQFTPNFRAPELK-----FQSVKVKAAEDLMGT 247
Db 256 HTMYTVHFYTGSHAASTESYPSE-----TPNSERGNVMSNTRYALENGVAVFATE--WGT 308
Qy 248 -----DSOLWFGYTOQ-----SHWOIFNGKNS-----RPFV----- 274
Db 309 SOASGCGGPFDEADVWIEFLNENNISWANNSLTN-KNEVSGAFTPFELGKSNATNLDPG 367
Qy 275 --HDYQP-EIFLT-----QPV---YSDLPW---DGKVRMIGMGAVHHSNGE 311
Db 368 PDHVWAPPELSLSGEYVRARIKGVNYEPIDRTKYTKVLWDFNDGTFKQ--GFGVNSDSPNK 425
Qy 312 SAKLSRSNRAYLMAGME-----WKNLTVMPRIWGR 342
Db 426 ELIAVDNENNTLKVSGLDVSDVDGNEFWANARLSANGWGK 466
```

Search completed: November 30, 2001, 14:26:30
Job time: 567 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 14:27:05 ; Search time 28.02 Seconds
(without alignments)
578.367 Million cell updates/sec

Title: US-09-787-083-8

Perfect score: 2363

Sequence: 1 MKVSLSTLTLSILPCFAILA.....YNHEATSGVGLMLNDWMGL 442

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	373	15.8	1 P1_KLEPN	P37446 klebsiella
2	358	15.2	1 P1_SALTY	P37442 salmonella
3	357	15.1	1 P1_ECOLI	P00631 escherichia
4	321.5	13.6	1 P1_PROVU	P37447 proteus vul
5	108.5	4.6	1 GUN_BACS1	P06564 bacillus sp
6	98.5	4.2	1 ANDM_YEAST	P15274 saccharomyc
7	95	4.0	1 YJH0_YEAST	P40361 saccharomyc
8	94.5	4.0	1 GUN_PAEPO	P23548 paenibacilli
9	94.5	4.0	1 BCK1_YEAST	Q01389 saccharomyc
10	93.5	4.0	1 SRK6_BRAOL	Q09092 brassica ol
11	92.5	3.9	1 VG14_BPML5	Q05220 mycobacteri
12	92.5	3.9	1 BAC2_MOUSE	P97303 mus musculus
13	92	3.9	1 YL70_ARCFU	O28112 archaeoglob
14	92	3.9	1 R114_HUMAN	P48552 homo sapien
15	91	3.9	1 CATA_DICDI	O77229 dictyostell
16	91	3.9	1 YP72_CAEEL	Q09221 caenorhabdi
17	90.5	3.8	1 INVO_MOUSE	P48997 mus musculus
18	89.5	3.8	1 PBP4_STRPN	Q04707 streptococ
19	89.5	3.8	1 Y075_MYCGE	P47321 mycoplasma
20	89.5	3.8	1 MET_MOUSE	P16056 mus musculus
21	89	3.8	1 NPPE_BACAM	P06832 bacillus am
22	88.5	3.7	1 DCUP_BACSU	P32395 bacillus su
23	88.5	3.7	1 YJHT_HAEIN	P44544 haemophilus
24	88.5	3.7	1 CD44_RAT	P26051 rattus norv
25	88.5	3.7	1 AM1_ALTHA	P29957 alteromonas
26	88.5	3.7	1 GUN3_BACS4	P19570 bacillus sp
27	88.5	3.7	1 SABC_STRSL	Q55242 streptococ
28	88	3.7	1 MURE_CHLTR	O84271 chlamydia t
29	88	3.7	1 DTXH_CORBE	P00589 corynebaph
30	87.5	3.7	1 SEIL_HUMAN	Q9ubv2 homo sapien
31	87	3.7	1 LPFC_SALTY	P43662 salmonella
32	87	3.7	1 DP3A_BACSU	O34623 bacillus su
33	87	3.7	1 A2MG_HUMAN	P01023 homo sapien

34	86	3.6	491	1	TY3H_PHASP	P11982 phasianidae
35	86	3.6	520	1	YEAE_SCHPO	O14079 schizosacch
36	86	3.6	666	1	PD14_RAT	O88807 rattus norv
37	86	3.6	828	1	BGAL_BRAOL	P49676 brassica ol
38	86	3.6	882	1	HSS1_RAT	Q02353 rattus norv
39	86	3.6	917	1	YGJ3_YEAST	P53148 saccharomyc
40	86	3.6	1087	1	XYNX_CLOTM	P38535 clostridium
41	86	3.6	1131	1	PM11_DUNBI	P54211 dunaliella
42	85.5	3.6	790	1	SEIL_MOUSE	Q92926 mus musculu
43	85.5	3.6	1390	1	MET_HUMAN	P08581 homo sapien
44	85.5	3.6	1840	1	SUIS_RAT	P23739 rattus norv
45	85	3.6	525	1	MP11_RAT	P48965 rattus norv

ALIGNMENTS

```

RESULT 1
PAL_KLEPN
ID PAL_KLEPN STANDARD; PRT; 286 AA.
AC P37446;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DETERGENT-RESISTANT
DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
GN PLDA.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94131966; PubMed=8300539;
RA Brok R.G.P.M., Brinkman E., van Bostel R., Bekkers A.C.A.P.,
RA Verheij H.M., Tomassen J.;
RT "Molecular characterization of enterobacterial plidA genes encoding
RT outer membrane phospholipase A.";
RL J. Bacteriol. 176:861-870(1994).
CC -|- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
CC -|- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -|- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -|- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
CC -|- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
CC LOCATED THERE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X76901; CAA54223.1; -
CC PIR; B36971; B36971.
CC PIR; S40129; S40129.
CC InterPro: IPR003187; PLAL.
CC Pfam: PF02253; PLAL; 1.
CC Hydrolase; Lipid degradation; Outer membrane; Signal; Calcium.
CC SIGNAL 1 20
CC CHAIN 21 286 PHOSPHOLIPASE A1.
CC ACT_SITE 161 161 BY SIMILARITY.
CC SEQUENCE 286 AA; 32544 MW; 3E39F863085108A3 CRC64;

```

Query Match 15.8%; Score 373; DB 1; Length 286;
Best Local Similarity 42.1%; Pred. No. 7.4e-23;
Matches 90; Conservative 27; Mismatches 87; Indels 10; Gaps 5;

[illegible]

```

RESULT 3
PAL_ECOLI
ID PAL_ECOLI STANDARD; PRT; 289 AA.
AC P00631;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DETERGENT-RESISTANT
DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
GN PLDA OR B3821 OR Z5342 OR ECS4751.
OS Escherichia coli, and
OS Escherichia coli 0157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=562, 83334;
[1]
SEQUENCE FROM N.A.
MEDLINE=85157492; PubMed=6397464;
Homma H., Kobayashi T., Chiba N., Karasawa K., Mizushima H., Kudo I.,
Ra Inoue K., Ikeda H., Sekiguchi M., Nojima S.;
RT "The DNA sequence encoding plda gene, the structural gene for
RT detergent-resistant phospholipase A of E. coli.";
RL J. Biochem. 96:1655-1664(1984).
[2]
SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=92358234; PubMed=1379743;
Ra Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT "Analysis of the Escherichia coli genome: DNA sequence of the region

```

RI from 04.5 to 06.5 minutes.
 RL Science 257:771-778(1992).
 [3]
 RN REVISION TO 14-15.
 RP
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).

Gregor, C. P., Bury, R. W., Karpatscher, H. H., & Osoeden, H. H. (1997).
 MAU B. Y., SHAO Y.;
 "The complete genome sequence of *Escherichia coli* K-12.";
 Science 277:1453-1474 (1997).
 [4]
 SEQUENCE FROM N.A.
 RC STRAIN=0157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G., Evans P.S., Gregor J., Karpatscher H.A.,
 RA Rose D.J., Mayhew G.F., Boutin A., Shao Y., Miller L.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamowitz K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";

WEICH K.A., BLANCHET F.A.,
"Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7." ;

RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / RIMD 0509952;
RC MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP SEQUENCE OF 30-289 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=85003590; PubMed=6383820;
RA de Geus P., Verheij H.M., Riegman N.H., Hoekstra W.P.M., de Haas G.H.;
RT "The pro- and mature forms of the E. coli K-12 outer membrane
RT phospholipase A are identical.";
RL EMBO J. 3:1799-1802(1984).
RN [7]
RP SEQUENCE OF 174-289 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=87115164; PubMed=3027506;
RA Irino N., Nakayama K., Nakayama H.;
RT "The recQ gene of Escherichia coli K12: primary structure and
RT evidence for SOS regulation.";
RL Mol. Gen. Genet. 205:298-304(1986).
RN [8]
RP MUTAGENESIS OF SER-172.
RX MEDLINE=94131966; PubMed=8300539;
RA Brok R.G.P.M., Brinkman E., van Bostel R., Bekkers A.C.A.P.,
RA Verheij H.M., Tommassen J.;
RT "Molecular characterization of enterobacterial plda genes encoding
RT outer membrane phospholipase A";
RL J. Bacteriol. 176:861-870(1994).
RN [9]
RP ACTIVE SITE SER-164, AND PARTIAL SEQUENCE.
RX MEDLINE=91249806; PubMed=2040286;
RA Horrevorts A.J.G., Verheij H.M., de Haas G.H.;
RT "Inactivation of Escherichia coli outer-membrane phospholipase A by
RT the affinity label hexadecanesulfonyl fluoride. Evidence for an
RT active-site serine.";
RL Eur. J. Biochem. 198:247-253(1991).
CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
CC REQUIRED FOR EFFICIENT SECRETION OF BACTERIOCINS, SEEMS TO BE
CC DORMANT IN NORMAL GROWING CELLS.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
CC LOCATED THERE.
CC -1- INDUCTION: BY MEMBRANE DAMAGING, FOR EXAMPLE, BY PHAGE-INDUCED
CC LYSIS OR TEMPERATURE SHOCK.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X02143; CAA26081.1; -
DR EMBL; M87049; AAC67617.1; -
DR EMBL; AE000458; AAC76824.1; -
DR EMBL; AE005613; AAG59017.1; -
DR EMBL; AF002567; BAB38174.1; -
DR EMBL; M30198; AAA24516.1; -
DR PIR; A00771; PSECA.

DR PIR; A22133; PSECAL.
DR PIR; S30711; S30711.
DR EcoGene; EG10738; plga.
DR InterPro; IPR003187; PLAI.
DR Pfam; PF02253; PLAI; 1.
KW Hydrolase; Lipid degradation; Outer membrane; Signal; Calcium;
FT Complete proteome.
FT SIGNAL 1 20 PHOSPHOLIPASE A1.
FT CHAIN 21 289
FT ACT_SITE 164 164
FT MUTAGEN 172 172
FT CONFLICT 14 15 S->F: INACTIVE PROTEIN.
FT CONFLICT 30 33 LP -> FA (IN REF. 2).
FT SEQUENCE 289 AA; 33163 MW; A688AD32AA60F218 CRC64;
Query Match 15.1%; Score 357; DB 1; Length 289;
Best Local Similarity 39.3%; Pred. No. 1.5e-21;
Matches 84; Conservative 29; Mismatches 91; Indels 10; Gaps 4;
QY 226 RAPELKFOVSVKVKAAEDLWGTDSLWFGYTOQSHWQIENGKNSRPRVRVHDYQPEIFLTQ 285
DB 83 RDEVAFOQLSLAPLWRGILGPNVSLGASTQKSWQLSNSESSPFRETNYEPQLFLGF 142
QY 286 PV-YSLPMDGKVRMIGMCAVHHNGESAKLSRSNRAYLMAGMKNLVTMPRMGRIF 344
DB 143 ATDYREFAGM--TLRDVEMGVNHDNCRSDPTSRSNRLYTRLMAENGWNLVEVKPWYV- 199
QY 345 KEGSGQPDNDPILDYGYGVRFYLYOLENKSNIQSVRYNPRSGKALQLDYVYPLGK 404
DB 200 -----GNTDDNDPITKMGYQKIGYHL-GDAVLAKGQYNWNTGYGGAELGSLSPITK 253
QY 405 GISGVIQIFOGYGOSLIDYNHEATSFVGVLMD 438
DB 254 HVRLYQVYSGYGESLIDYNFNQTRVGVGVMLND 287
RESULT 4
PAL_PROVU STANDARD; PRT; 289 AA.
AC P37447;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DETERGENT-RESISTANT
DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
GN PLDA.
OS Proteus vulgaris.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Proteus.
OX NCBI_TaxID=585;
RP [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=94131966; PubMed=8300539;
RA Brok R.G.P.M., Brinkman E., van Bostel R., Bekkers A.C.A.P.,
RA Verheij H.M., Tommassen J.;
RT "Molecular characterization of enterobacterial plda genes encoding
RT outer membrane phospholipase A";
RL J. Bacteriol. 176:861-870(1994).
CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
CC LOCATED THERE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X02143; CAA26081.1; -
DR EMBL; M87049; AAC67617.1; -
DR EMBL; AE000458; AAC76824.1; -
DR EMBL; AE005613; AAG59017.1; -
DR EMBL; AF002567; BAB38174.1; -
DR EMBL; M30198; AAA24516.1; -
DR PIR; A00771; PSECA.

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X76902; CAA54224.1; -;
DR PIR; C36971; C36971.
DR PIR; S40130; S40130.
DR InterPro; IPR003187; PLA1.
DR Pfam; PF02253; PLA1; 1.
KW Hydrolase; Lipid degradation; Outer membrane; Signal; Calcium.
FT SIGNAL 20 BY SIMILARITY.
FT CHAIN 21 289 PHOSPHOLIPASE A1.
FT ACT_SITE 164 164 BY SIMILARITY.
SQ SEQUENCE 289 AA; 32944 MW; D75516CFEB406997 CRC64;

Query Match 13.6%; Score 321.5; DB 1; Length 289;
Best Local Similarity 33.3%; Pred. No. 1e-18;
Matches 87; Conservative 42; Mismatches 115; Indels 17; Gaps 7;

Qy 179 LDRNNTPLWSSRPHPMVLPIFMHGKPNRSPNTSPSHEAKOFTPNFEFRAPELKFQVSVKV 238
Db 43 LQEHDP-FTLYPYESNLLTYT-----TSDLNKAIESYNWSDNA-NKDEVKFOLSLAF 95
Qy 239 KAAEDLWGTDSLWFGYQQSHWQIFNGKNSRPFRVHDYQPEIFLTQPV-YSDLPLWDGKV 297
Db 96 PLWRGILGDSLLGASYTQRSWQLSNTGESAPFETNYEPQLFLGFATDYSVGDW--TL 153
Qy 298 RMIGMGAVHHNGESAKLSRSNNRAYLMAGMEWKNLTVMPTRWGRIFKEGSGQDDNPD 357
Db 154 RDAEFGYNHQNSRDPTRSRNQLIYSLMAQNGNWLVEVKPVIYI-----GDTSDNKN 207
Qy 358 ILDYYGYGDVRFYOLEKNSNIGTVRNPNSRGKALQIDYVPLGKISGYFQIFQGYG 417
Db 208 ITKYMGGYQLKIGYQL-GEAVLSAKGQYNWNTGYGAELGVSGYPTTKHVRFTYQVSYGY 266
Qy 418 QSLIDYNEHATSFYGVGLMLND 438
Db 267 ESLIDYDFNQIRVNGVGLND 287

RESULT 5
GUN_BACS1 STANDARD; PRT; 800 AA.
AC P06564;
DT 01-JAN-1998 (Rel. 06, Created)
DT 01-JAN-1998 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (ALKALINE CELLULASE).
OS Bacillus sp. (strain 1139).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1411;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87085443; PubMed=30989509;
RA Fukumori F., Kudo T., Narahashi Y., Horikoshi K.;
RT "Molecular cloning and nucleotide sequence of the alkaline cellulase
gene from the alkalophilic Bacillus sp. strain 1139.";
RL J. Gen. Microbiol. 132:2329-2335(1986).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
LINKAGES IN CELLULOSE.
CC -1- MISCELLANEOUS: ALKALOPHILIC BACILLUS SP., STRAIN 1139, PRODUCES
ONE ALKALINE CELLULASE (PH OPTIMUM 9.0). THIS STRAIN IS NOT A TRUE
CELLULOLYTIC MICRO-ORGANISM BECAUSE THE ENZYME IS UNABLE TO
HYDROLYSE NATIVE CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M15743; AAA22305.1; -;
DR EMBL; D00066; BAA00045.1; -;
DR PIR; A29003; A29003.
DR InterPro; IPR001547; Glyco_hydro_F5.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 800 ENDOGLUCANASE.
FT ACT_SITE 190 190 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 305 305 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 800 AA; 88602 MW; 7CCA4D7B6DAD5CF CRC64;

Query Match 4.6%; Score 108.5; DB 1; Length 800;
Best Local Similarity 19.9%; Pred. No. 0.6;
Matches 96; Conservative 54; Mismatches 159; Indels 173; Gaps 25;

Qy 8 LTLSELPCFAILAIQAOAVPNPVAFVDEVRSND---LGQDNELPIDVQSATOSASTDT 64
Db 10 LISSILILVLLSL-----FPTALAAEGNTRDNFKLLGNDNVKRPSEAGALQLEVDG 64
Qy 65 ANPL-DEHEPELY-----TTALENKTMLNCSAL-----NODIMRLACY-DTLVHGET 110
Db 65 QMTLVQDHGKIQLRGSMTHGLWFFPEILNDNAYKALANDWESNNIRLAMYVGNGYASN 124
Qy 111 PAVIKTKRSIRLDEPI-----WQT-----IKGKPOVV 137
Db 135 PELKSRVVIKIDIAIENDMYIVDVHWHVHAGDPDPVYAGAEDEFRIALYNNPHII 184
Qy 138 YQETTP-----IFLMGNEKGMITKKDAKOLEYAAKQFTPLSLSFOLDNRN---NTP 186
Db 185 YELANEPSSNNGGAGIPNNEEGNAVK-----EVADPIVEMLRDSDGNADDNIIIVGSPN 239
Qy 187 WSSRP-----HNPMYVLPF--MHGKPNRS--PNTPSHEAKQFTPNFEFRAPELKF 232
Db 240 WSQRPDLAADNPIDDHHTMYTVHFYTGSHAASTESYPETPNSEGRGNYSNTRYA-----L 295
Qy 233 QVSKVKAEDLWGT-----DSLWFGYTOQ-----SHWQIFNGKNS-----R 270
Db 296 ENGVAVFATE--WGTSQANGDGGPYFDEADVWIEFLNENNISWANWSLTN-KNEVSGAFT 352
Qy 271 PFRVH-----DYQPELFLTPQVYSIDLPMW--- 293
Db 353 PFELGKSNATSLDPGPDQVWVPEELSLSGEVVRARIKGVNVEP---IDRTKYTKVLWDFN 409
Qy 294 DGKVRMIGMGAVHHSNGES-----AKLSRSNNRAYLMAGMEWKNLTVMPTRIW 340
Db 410 DGTQKGFV-----NGDSPVEDVVIENAGALKUSGLDASNDVSEGNYNANARLSADGW 463
Qy 341 GR 342
Db 464 GK 465

RESULT 6
ADML_YEAST STANDARD; PRT; 810 AA.
AC P15274;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE AMP DEAMINASE (EC 3.5.4.6) (MYOADENYLATE DEAMINASE).
GN AMD1 OR AMD OR YML035C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

RESULT	YJHO_YEAST	STANDARD;	PRT;	888 AA.
ID	YJHO_YEAST			
AC	P40361;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	HYPOTHEITICAL 104.3 KDA PROTEIN IN SMC3-MRPL8 INTERGENIC REGION.			
GN	YJL070C OR J1095 OR HRD888.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
NCBI_TaxID=4932;				
ON	[1]			
RP	SEQUENCE FROM N. A.			
RC	STRAIN=S288C;			
RA	MEDLINE=95282514; PubMed=7762302;			
RX	Vandenbol M., Durand P., Dion C., Portetelle D., Hilger F.;			
RT	"Sequence of a 17.1 kb DNA fragment from chromosome X of			
RT	Saccharomyces cerevisiae includes the mitochondrial ribosomal protein			
RL	L8.;"			
RL	Yeast 11:57-60(1995).			
RN	[2]			
RP	SEQUENCE FROM N. A.			
RC	STRAIN=S288C;			
RA	Sor F. J.;			
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: BELONGS TO THE ADENOSINE AND AMP DEAMINASES FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announcement/			
CC	or send an email to license@lsb-sib.ch).			
CC	EMBL; Z34288; CAA84052.1; -;			
CC	EMBL; Z49345; CAA89362.1; -;			
DR	EMBL; Z49344; CAA89361.1; -;			
DR	EMBL; X88851; CAA61309.1; -;			
DR	PIR; S47120; S47120.			
DR	SGD; S0003606; YJL070C.			
DR	InterPro; IPR001365; A_deaminase.			
DR	Pfam; PF00962; A_deaminase; 1.			
DR	Hypothetical protein; Hydrolase.			
SK	SEQUENCE 888 AA; 104263 MW; DB31A0806224114D CRC64;			

	Query Match	4.08;	Score 95;	DB 1;	Length 888;
	Best Local Similarity	22.5%;	Pred. No. 8.4;		
	Matches	67;	Conservative	35;	Mismatches 126; Indels 70; Gaps 15;
Qy	128 QTIKGKPOVVYQE-----TTDPFIPLMGNEKMGLTKKDA-----KOLEYAANKQFTPLSLS	176 : : : :			
Dd	2 QAVERRPSLLFDEYQNYSVTPNETKNKEARVLSENDGDGSPSVLUKQEKISVDYDDMDMISLP	61 : : : :			
Qy	177 FDLDRN--NTPLW---SSRPHPNMVLPPIPMHGKPNR-----SNPTPISH-----EAKQF	220 : : : :			
Dd	62 TEFDRQMVLSPMFDFLEBENKDIDPLPSVSHHYNGESGSFVSYPSPNLUKTGTEYKDL	121 : : : :			
Qy	221 TPNFRAPELKFQSVKVKAADLWG-----TSDSLWFGYTQ---QSHWQIFNGKNSRPF	273 : : : :			
Dd	122 FINFP---ELVSQRKRXYIAASKODGISNIKNDEFKFWFLPYKPLPKFWREDDK----	R 173 : : : :			
Qy	274 VHDYQEIFLTQPYYSDLPWDGKYVMTCGMGAHVHHSNGESAKLRSRWRAYILMAWMEMKNL	333 : : : :			
Dd	174 FQD-----PSDSLNDGGDTGTGAATPHRHGYGPYSFYTDTHYYYYTSCLKCKGN	224 : : : :			
Qy	334 TVMPRIINGIRFKEGSGSQPDNDPIDLYYGVDGVRFYLQLENKSINIGTVRYNPSPSK	391 : : : :			
Dd	225 IKVP-YTGEYF-----DLEDY----KKOYIYHLSNSOBNTONPL--SPWSX	263 : : : :			


```
FT CARBOHYD 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 389 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 442 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 849 AA; 97231 MW; 7E156059EDDF4370 CRC64;

Query Match
Best Local Similarity 21.9%; Pred. No. 10; Length 849;
Matches 78; Conservative 50; Mismatches 107; Indels 121; Gaps 24;

QY 105 LVHGETPAVITKRSIRLDETQITKQPVVYQETDPIFLMG-----NEK---GMLTK 157
DQ 24 LIHPALSIYINT-----LSSTESLTISNNKTLV---SPGISPEVGFRTNSRWYLCGMVK 75
QY 158 K-DAKOLEVAAKQFTPLSLSPD-----LDNRNTPWSSRPHNPMYVLPFIMHGK 205
DQ 76 KVSDDTYVWVNRDPLNSAIGTLKISNNVLLDHSNKPVMWTN-----LTRGN 125
QY 206 PNRSP-----NTPSHEAKQFT-----PNEFRAPELKFQSVKVKAAEDLW 245
DQ 126 -ERSPVAELLANGFNVRDSSNNDASEYLVQSFDPDITLLPENKLGYNLKT----- 177
QY 246 GTDSDLVGYTQSHQIFNGKNSRFRVHDYQ-----PEIFLQO---PVYSDLPWDGK 296
DQ 178 GLNREF-----TSWRSDDPSSGNF---SYKLETQSLPEFYLSENFPHRSRGPWNG- 226
QY 297 VRMICMGAVH-----HSGESAKLSRSNNRAY-----LMAGMEWKNLTVMF--RI 339
DQ 227 IFSGIPEDQKLSYVWYFNFIENNEVATFTNTNSFVRLTLISEGFQRLTWTPYSIRI 286
QY 340 WGRIFKEGSGQPDNDPILDY-----XYGVDVRFYQLENKSNISGTVR--YNPRS 389
DQ 287 NNRFW-----SSPVD-PQCDTYMGCPAYCDV-----NTSPVCNIOGPNRN 329

RESULT 11
VG14_BPML5
ID VG14_BPML5 STANDARD; PRT; 486 AA.
AC Q05220;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE GENE 14 PROTEIN (GP14).
GN 14.
OS Mycobacteriophage L5.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
OX NCBI_TaxID=31757;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93211282; PubMed=8459766;
RA Hatfull G.F., Sarkis G.J.;
RT "DNA sequence, structure and gene expression of mycobacteriophage L5:
RT a phage system for mycobacterial genetics.";
RL Mol. Microbiol. 7:395-405(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z18946; CAA79390.1;
DR PIR; S30959; S30959
SQ SEQUENCE 486 AA; 53757 MW; 4BF3734AC6B11BF5 CRC64;
```

```
Query Match
Best Local Similarity 3.9%; Score 92.5; DB 1; Length 486;
Matches 99; Conservative 69; Mismatches 173; Indels 131; Gaps 24;
```

```
QY 21 IQQAQA-VPNVAFVDEVRSEN-----DLGQDNELPIDVQSATQASASTDTANPLDHEPE 74
DQ 56 MOQLLAHVGYPRLYVDSVAERQAVGFRGLGDADEAEELQWQWQ-----ANNLDIEAPL 109
QY 75 LYTTALENKTMLNCSA-----LNQDIMRLACVDTLVHGE-TPAVIKTKRSIRLD 123
DQ 110 GYTDAYVHGSRSPITISKPPDQLDLGWDQNVPIRVE-PPTRMHAEIDPRINRVSKAIR-- 166
QY 124 ETIWTQIKGQVQVYQETDPIFLMGNEKGMLTKKDAKQLEVAAKQFTPL-SLSFDLDRN 182
DQ 167 -----VAYDKE-----GNE-----TQAATLYTPMETIGFRADG 195
QY 183 NTPWSSRPHNPMYVLPFIMHGKPNRSNPTPSHEAKQFTPEFRAPELKFQSVKVKAAE 242
DQ 156 EWAENFNVPHG-LGVVPV--PLPNRTRLSLDYGTSEITPELRSMTDAAARILMLQATA 252
QY 243 DLWGTDSDLNF-----GYTQSHQIFNGKNSRFRVHDYQPEI-----F 282
DQ 253 ELMGVPPQRLIFGIKPEEIGVDSETGQTLFDAYLARILAFEDAEGKIQOFSAAELANFTNA 312
QY 283 LTQ-----PVYSDLPWDGKVRMIGMGAVHSGNES-----AKLSRSWNRAYLMAGMEWKN 332
DQ 313 LDQIAKQVAAITGLP-----PQYLSTAADNPASAEAIRAAESRLIKKVERKNLMFGCAHEE 368
QY 333 LTVMPRIWGRIFKEGSGQPDNDPILDYGYGVDVRFYQLENKSNISGTVRYNPRSGK 392
DQ 369 AM---RIAYRIMKGG-----DVPPDML-----RME-----TVWRDPSTPTY 401
QY 393 ALOLD---YVYPLGKI--SGYFQIFQGYG-----QSLIDYNHEATSFVGVLG 435
DQ 402 AAKADAATKLYGNGGQGVIPRERARIDMGYSVKEREEMRRWDEEAAAGLGL 453

RESULT 12
BAC2_MOUSE
ID BAC2_MOUSE STANDARD; PRT; 716 AA.
AC P97303;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSCRIPTION REGULATOR PROTEIN BACH2 (BTB AND CNC HOMOLOG 2).
GN BACH2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=97042438; PubMed=8887638;
RA Oyake T., Itoh K., Motohashi H., Hayashi N., Hoshino H., Nishizawa M.,
RA Yamamoto M., Igarashi K.;
RT "Bach proteins belong to a novel family of BTB-basic leucine zipper
RT transcription factors that interact with Mafk and regulate
RT transcription through the NF-E2 site.";
RL Mol. Cell. Biol. 16:6083-6095(1996).
CC -|- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT ACTS AS REPRESSOR OR
CC ACTIVATOR. BINDS, IN-VITRO, TO NF-E2 BINDING SITES. PLAY IMPORTANT
CC ROLES IN COORDINATING TRANSCRIPTION ACTIVATION AND REPRESSION BY
CC MAFK.
CC -|- SUBUNIT: HETERO-DIMER OF BACH2 AND MAFK.
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -|- TISSUE SPECIFICITY: EXPRESSION RESTRICTED TO MONOCYTES AND
CC NEURONAL CELLS.
CC -|- SIMILARITY: BELONGS TO THE BZIP FAMILY. CNC SUBFAMILY.
CC -|- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; D86604; BAA13138.1; -
DR HSSP; P05412; IFOS.
DR MGD; MGI:894679; Bach2.
DR InterPro; IPR000210; BTB_POZ.
DR InterPro; IPR001871; bZIP.
DR Pfam; PF00651; BTB; 1.
DR Pfam; PF00170; bZIP; 1.
DR SMART; SM00338; BRL2; 1.
DR SMART; SM00225; BTB; 1.
DR PROSITE; PS50097; BTB; 1.
DR PROSITE; PS00036; bZIP_BASIC; 1.
KW Transcription regulation; Activator; Repressor; DNA-binding;
KW Nuclear protein.
FT DOMAIN 37 103 BTB.
FT DOMAIN 162 168 POLY-GLU.
FT DNA_BIND 527 542 BASIC MOTIF.
FT DOMAIN 550 572 LEUCINE-ZIPPER.
SQ SEQUENCE 716 AA; 78935 MW; 9132B3731AE24333 CRC64;

Query Match 3.9%; Score 92.5; DB 1; Length 716;
Best Local Similarity 22.8%; Pred. No. 10; Mismatches 127; Indels 149; Gaps 21;
Matches 93; Conservative 39;

QY 38 RSENDLG-----QNELPIDVQSATSASTD-----TANPLDEH-----EPE 74
DB 151 RPQEDHGNSAGEEEEETMDSETARMACATDMLPDPISFEATALPVAKEEALLPESE 210
QY 75 LYTALENKTMINSALNQ-----DIMRLAC-----YDTLVHG----- 108
DB 211 VPTDKENSEK-----GALTQYPRYKYLQACTKNYSAPSHSGTGFASFSDSPGNSLK 266
QY 109 -----ETPAVKTKRSIRL-----DETWTQIKGPOVYQETDPI----- 145
DB 267 PGLPMGQIKSEPPSEETEEETLCSGDET-----DIKDRPGDVMDRKQPSPARTPTRT 323
QY 146 -----FLMNEKGMLTKDKAKOLEAAKQFTPLSLF----- 177
DB 324 GAACLDRSVSSPCLSLFGITGV-----ESTGLPSTSQ--PLVRSSACPPNKGISQ 377
QY 178 -DLDRNPTPLWSS--RPHNPMYVLPFIHMGKPNRSP-----NTPSHEAKQFTPNF 225
DB 378 GDLKTDYTPLAGNYGQPHVQGVQDVSNFAMGSPLRGPGPETLCEFSQSCQAGRELATEH 437
QY 226 RAPELK-----FQSVKVKAAEDLNGTSDLMFGYTTQOSHWFIFNGKNSRPFVRHDYQPE 280
DB 438 QEPGLMGDMYQVVRQIK-CEQSYGTNSSDESG-----SFSEADSESCPVQDRGOE 488
QY 281 IFLTQPV--YSDLPWDGKVRMIGMGAVHNSGESAKL-----SRSWNR 321
DB 489 VKLPFPVDOITDLPNDQFMIMK---HKLSEQLEFTHDIRRSKNR 533

RESULT 13
YL70_ARCFU STANDARD; PRT; 324 AA.
AC Q28112.
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN AF2170.
GN AF2170.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
```

```
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Cocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -!- SIMILARITY: CONTAINS 2 KELCH REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; AE000955; AAB89093.1; -.
DR TIGR; AF2170; -.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF01344; Kelch; 2.
KW Hypothetical protein; Repeat; Complete proteome.
FT REPEAT 229 276 KELCH 1.
FT REPEAT 277 323 KELCH 2.
SQ SEQUENCE 324 AA; 36025 MW; C445388CFEB96E45 CRC64;

Query Match 3.9%; Score 92; DB 1; Length 324;
Best Local Similarity 20.1%; Pred. No. 3.7; Mismatches 44; Indels 98; Gaps 18;
Matches 54; Conservative 44;

QY 224 EFRAPELKQVSVKVKAAED-----LWGTSDSLMFGYTTQOSHWFIFNGKNSRPFVRH 275
DB 63 EHFPPN-RLIEI---LLSSEDFSQREAANVWDGEILFQGT-----VFENGKYSPTDQIL 112
QY 276 DYQPEIFLTQPVYSDLP-----W-DGKVRMI-----GMGAVHNSGESAKLSRS 318
DB 113 SNPKLERLRVLNASLPHPTSDVAAVWGDVSRVYIFLNSRCEVAFYPSNESFAKLQVS 172
QY 319 -----W--NRAYL-----MAGMEKNLTVMPIRIGRIFKEGSGSQ 351
DB 173 CPTEHPGGCVHSVWVYGGKAYFFCGEGVASFDPMGGFKW--IAFTDRVWVRAATVADG-- 228
QY 352 PDNPDILDYGYGVDVRFYQLENKSNISGT-----VRYNPRSGKQALQDYVYPLGKG-- 405
DB 229 -----YIFAIGSGSGIAETKDEIRFPNPKTGE-LCEMRKLPVARGQA 270
QY 406 --ISG-YFOIF-----QGYQSLIDYNH 425
DB 271 VAVGGEYIYIFGGYTKDGYANEILRYDY 298

RESULT 14
RI14_HUMAN STANDARD; PRT; 1158 AA.
AC P48552;
DT 01-FEB-1996 (Rel. 33, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NUCLEAR FACTOR RI140 (NUCLEAR RECEPTOR INTERACTING PROTEIN 1).
GN NR1P1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
```

[illegible]

```

QY 157 KDAKOLEYAAKQFTPLSLSFDLDRNNTPLWSSRPHPNMYVLPFIEMHGKPNRSPNTPPSHE 216
:|: | | : | | : | | : | | : | | : | | : | | : | | :
Db 106 ERDPR--GEAVKFYTEEG-NFDMVGNNTPVFFIRDPK---FPDFIHTQ-KRNPQTNCCKD 158
QY 217 AKOFTPNEPRAPELKTOVSVKYKAAEDLWGTDSDLWFGYTQQSHWQIFNGKNSRPFRV 274
:|: | | : | | : | | : | | : | | : | | : | | :
Db 159 PNMFWDFLGOTPESTHOVSILFSDR---GTPKSYRHHMGFSSHTLKFVNAQCKPYWV 212

```

Search completed: November 30, 2001, 14:27:06
Job time: 553 sec

This Page Blank (uspto)